

NONLINEAR LARGE DEVIATIONS

SOURAV CHATTERJEE* AND AMIR DEMBO†

ABSTRACT. We present a general technique for computing large deviations of nonlinear functions of independent Bernoulli random variables. The method is applied to compute the large deviation rate functions for subgraph counts in sparse random graphs. Previous technology, based on Szemerédi's regularity lemma, works only for dense graphs. Applications are also made to exponential random graphs and three-term arithmetic progressions in random sets of integers.

1. INTRODUCTION

1.1. A motivating example. Let $G(N, p)$ be the Erdős–Rényi random graph on N vertices with edge probability p , that is, the classical model where any two vertices are connected by an edge with probability p , independent of all else. Let T denote the number of triangles in this graph. It has been an open question in the random graph literature for a long time [20] to determine the behavior of the upper tail of T , that is, probabilities of the type $\mathbb{P}(T \geq (1 + \delta)\mathbb{E}(T))$. The main difficulty with this problem, and the reason why it may be appealing to a probabilist, is that the standard tools from concentration of measure and other probability inequalities do not seem to work so well in this setting, in spite of the fact that the number of triangles in an Erdős–Rényi graph is simply a degree three polynomial of independent Bernoulli random variables.

After a series of successively improving suboptimal results by many authors over many years, a big advance was made by Kim and Vu [25] and simultaneously by Janson et al. [19] in 2004 who showed that if $p \geq N^{-1} \log N$, then

$$\exp(-c_1(\delta)N^2p^2 \log(1/p)) \leq \mathbb{P}(T \geq (1 + \delta)\mathbb{E}(T)) \leq \exp(-c_2(\delta)N^2p^2),$$

where $c_1(\delta)$ and $c_2(\delta)$ are constants depending on δ only.

Several years later, the logarithmic discrepancy between the exponents on the two sides was removed by Chatterjee [10] and independently by DeMarco and Kahn [15, 16], where it was shown that when $p \geq N^{-1} \log N$,

$$\begin{aligned} \exp(-c_1(\delta)N^2p^2 \log(1/p)) &\leq \mathbb{P}(T \geq (1 + \delta)\mathbb{E}(T)) \\ &\leq \exp(-c_2(\delta)N^2p^2 \log(1/p)). \end{aligned}$$

This still left open the question of determining the dependence of the exponent on δ . When p is fixed and N tends to infinity, the problem was solved in 2011 by Chatterjee and Varadhan [13], confirming a conjecture from an unpublished manuscript of Bolthausen, Comets and Dembo [3]. In [13], it was shown that for fixed $p \in (0, 1)$ and $\delta > 0$,

$$\mathbb{P}(T \geq (1 + \delta)\mathbb{E}(T)) = \exp(-c(\delta, p)N^2(1 - o(1))) \quad (1.1)$$

2010 *Mathematics Subject Classification.* 60F10, 05C80, 60C05, 05A20.

Key words and phrases. Large deviations, sparse random graphs, regularity lemma, arithmetic progressions, exponential random graph models, concentration of measure.

* research partially supported by NSF grant DMS-1441513.

† research partially supported by NSF grant DMS-1106627.

as $N \rightarrow \infty$, where

$$c(\delta, p) = \frac{1}{2} \inf_f \{I_p(f) : T(f) \geq (1 + \delta)p^3\}, \quad (1.2)$$

where $f : [0, 1]^2 \rightarrow [0, 1]$ is any Lebesgue measurable function that satisfies $f(x, y) = f(y, x)$ for all x and y ,

$$I_p(f) = \iint_{[0,1]^2} \left(f(x, y) \log \frac{f(x, y)}{p} + (1 - f(x, y)) \log \frac{1 - f(x, y)}{1 - p} \right) dx dy,$$

and

$$T(f) = \iiint_{[0,1]^3} f(x, y) f(y, z) f(z, x) dx dy dz.$$

Incidentally, the variational problem (1.2) has not yet yielded explicit solutions except in special ranges of δ and p [11, 13, 28, 29].

The above result was proved using Szemerédi's regularity lemma [31] from graph theory. A well known problem with Szemerédi's lemma is that it yields very poor quantitative bounds, which makes it virtually impossible to extend the arguments of [13] to the case where p is allowed to tend to 0 as $N \rightarrow \infty$. One can show (e.g. in [29]) that a weaker version of Szemerédi's lemma suffices for the proof in [13], which makes it possible to make the technique work when p tends to zero slower than a negative power of $\log N$, but it seems safe to bet that a Szemerédi type argument cannot help when p goes to zero like a negative power of N .

The last problem mentioned in the previous paragraph, namely, computing $c(\delta, p)$ in (1.1) when p goes to zero like $N^{-\alpha}$ for some $\alpha > 0$, was the original motivation for this paper. What we accomplish in this article is the following: We build a general machinery for tackling large deviations of certain class of nonlinear functions of independent Bernoulli random variables, which in particular circumvents the use of Szemerédi's lemma. Among other things, this approach yields a variational formula for $c(\delta, p)$, analogous to (1.2), which holds when $p = p(N) \rightarrow 0$ slower than $N^{-1/42}$. To see what its potential benefits may be, we note that after the first version of this paper was posted on arXiv, Lubetzky and Zhao [29] found a way to explicitly solve our variational problem when $N^{-1} \ll p \ll 1$. As a result of this additional exciting development, we now know that if $p(N) \rightarrow 0$ slower than $N^{-1/42}$, then

$$\mathbb{P}(T \geq (1 + \delta)\mathbb{E}(T)) = \exp\left(- (1 - o(1)) \min\left\{\frac{\delta^{2/3}}{2}, \frac{\delta}{3}\right\} N^2 p^2 \log \frac{1}{p}\right).$$

Therefore, the variational formula for $c(\delta, p)$ proved in this paper and its solution by Lubetzky and Zhao have completed the quest for understanding the behavior of the upper tail of triangle counts in Erdős–Rényi random graphs under the restriction that $p \gg N^{-1/42}$. It has been conjectured in [29] that the same formula should hold all the way down to $p \gg N^{-1/2}$. A strong evidence in favor of this conjecture is that the Lubetzky–Zhao solution of the variational formula for $c(\delta, p)$ holds whenever $p \gg N^{-1/2}$.

1.2. Goal of the paper. Suppose that $f : [0, 1]^n \rightarrow \mathbb{R}$ is a function with some amount of smoothness. Let p be a number in the open interval $(0, 1)$, and $Y = (Y_1, \dots, Y_n)$ be a vector of i.i.d. *Bernoulli*(p) random variables. For $u \in [0, 1]$ let

$$I_p(u) := u \log \frac{u}{p} + (1 - u) \log \frac{1 - u}{1 - p}, \quad (1.3)$$

and for each $x = (x_1, \dots, x_n) \in [0, 1]^n$, define

$$I_p(x) := \sum_{i=1}^n I_p(x_i). \quad (1.4)$$

For each $t \in \mathbb{R}$, define

$$\phi_p(t) := \inf\{I_p(x) : x \in [0, 1]^n \text{ such that } f(x) \geq tn\}. \quad (1.5)$$

Our goal is to investigate conditions under which the following “upper tail approximation” is valid:

$$\mathbb{P}(f(Y) \geq tn) = \exp(-\phi_p(t) + \text{lower order terms}). \quad (1.6)$$

Analogous statements may be similarly formulated if the Y_i ’s have some distributions other than Bernoulli.

It is easy to show that such an approximation is valid for all “effectively linear” maps, namely for continuous maps of the empirical measure of (Y_1, \dots, Y_n) . This fact forms the basis of a big part of modern large deviations theory, see [17] and the references therein. With some extra, often considerable, effort, one is able to extend this to “nearly continuous” maps of the empirical measure of (Y_1, \dots, Y_n) . One challenging example of the latter type, is the recent proof in [13] that the approximation holds for upper tails of subgraph counts in dense Erdős-Rényi random graphs, a result which was later generalized to random matrices [14] and exponential random graphs [12].

The proofs in [13, 14, 12] are, however, rather specialized to the random graph setting. The main tool in these papers is the regularity lemma of Szemerédi [31] and the graph limit theory of Lovász and coauthors [5, 6, 27] that builds on the regularity lemma. The unavailability of a suitable “sparse” version of Szemerédi’s lemma makes it impossible to extend the results of [13, 14, 12] to sparse graphs. Serious attempts have been made at formulating a sparse graph limit theory and sparse regularity lemmas [2, 4], but it is unlikely that these will provide the precision required for large deviations. The reason is that in all existing formulations, there is always some assumption about the regularity of the graph structure, and it may not be true that random graphs obey such regularity conditions in the large deviations regime. In graph theoretic terminology, the absence of a “counting lemma” for sparse graphs is the main impediment to extending a Szemerédi type argument to the sparse case.

More importantly, ideally one should not need to resort to specialized graph theoretic tools to prove an approximation as simple and basic as (1.6) for an f that may be as uncomplicated as a polynomial of degree three (e.g. number of triangles).

Our main objective here is to give a general error bound for the approximation (1.6) directly in terms of properties of the function f (as elaborated in the sequel), with an error bound small enough to allow extension of the aforementioned graph theoretic large deviation results to sparse random graphs. Incidentally, there are several notable results on upper bounds for tail probabilities for nonlinear functions of independent Bernoulli random variables. The bounded difference inequality [30] has been available for a long time. Improved inequalities were discovered by Talagrand [32], Łatała [26], Kim and Vu [24] and Vu [35]. However, all these methods seem to fall short of proving an approximation such as (1.6).

1.3. The main result. Our main result is Theorem 1.1 which gives a sufficient condition for the validity of the approximation (1.6). This sufficient condition may be roughly described as follows: The approximation (1.6) is valid when, in addition to some minor smoothness conditions on the function f , *the gradient vector $\nabla f(x) = (\partial f/\partial x_1, \dots, \partial f/\partial x_n)$ may be approximately encoded by $o(n)$ bits of information.* One may call this the “low complexity gradient” condition.

To illustrate this, consider the simple case of

$$f(x) = \sum_{i=1}^{n-1} x_i x_{i+1},$$

where the approximation (1.6) is not valid. Indeed, large deviation probabilities for this function are related to the one-dimensional Ising model, and easily shown to not satisfy (1.6). For this function, $\partial f / \partial x_i = x_{i-1} + x_{i+1}$ for $2 \leq i \leq n-1$, $\partial f / \partial x_1 = x_2$, and $\partial f / \partial x_n = x_{n-1}$, so clearly this gradient vector cannot be approximately encoded by $o(n)$ many bits; we effectively need to know all the x_i 's to encode the gradient vector of this function. On the other hand, if

$$f(x) = \frac{1}{n} \sum_{1 \leq i < j \leq n} x_i x_j,$$

as in the Currie-Weiss model, then for each i ,

$$\frac{\partial f}{\partial x_i} = \frac{1}{n} \sum_{j \neq i} x_j = -\frac{x_i}{n} + \frac{1}{n} \sum_{j=1}^n x_j.$$

Thus, the gradient vector is approximately encoded by the single quantity $n^{-1} \sum x_j$, so the “low complexity gradient” condition holds and the large deviation probabilities (which in this trivial case are covered by the general theory of large deviations), satisfy (1.6).

Unfortunately, although its content matches very well the preceding description, the actual statement of the theorem is somewhat messier, and requires some additional notation which we introduce next.

Let $\|f\|$ denote the supremum norm of $f : [0, 1]^n \rightarrow \mathbb{R}$. Suppose that $f : [0, 1]^n \rightarrow \mathbb{R}$ is twice continuously differentiable in $(0, 1)^n$, such that f and all its first and second order derivatives extend continuously to the boundary. For each i and j , let

$$f_i := \frac{\partial f}{\partial x_i} \quad \text{and} \quad f_{ij} := \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

Define

$$a := \|f\|, \quad b_i := \|f_i\| \quad \text{and} \quad c_{ij} := \|f_{ij}\|.$$

Given $\epsilon > 0$, let $\mathcal{D}(\epsilon)$ be a finite subset of \mathbb{R}^n such that for all $x \in \{0, 1\}^n$, there exists $d = (d_1, \dots, d_n) \in \mathcal{D}(\epsilon)$ such that

$$\sum_{i=1}^n (f_i(x) - d_i)^2 \leq n\epsilon^2. \tag{1.7}$$

The following theorem gives an error bound for the approximation (1.6) in terms of the quantities a , b_i , c_{ij} and the sizes of the sets $\mathcal{D}(\epsilon)$.

Theorem 1.1. *For f as above, $p \in (0, 1)$ and Y a vector of n i.i.d. Bernoulli(p) random variables, let ϕ_p be defined as in (1.5). Then, for any $\delta > 0$, $\epsilon > 0$ and $t \in \mathbb{R}$,*

$$\begin{aligned} \log \mathbb{P}(f(Y) \geq tn) &\leq -\phi_p(t - \delta) + \text{complexity term} \\ &\quad + \text{smoothness term}, \end{aligned}$$

where with $a, b, c_{ij}, \mathcal{D}(\epsilon)$ defined above, and $K := \phi_p(t)/n$,

$$\begin{aligned} \text{complexity term} &:= \frac{1}{4} \left(n \sum_{i=1}^n \beta_i^2 \right)^{1/2} \epsilon + 3n\epsilon + \log \left(\frac{4K \left(\frac{1}{n} \sum_{i=1}^n b_i^2 \right)^{1/2}}{\delta \epsilon} \right) \\ &\quad + \log |\mathcal{D}((\delta \epsilon)/(4K))|, \quad \text{and} \\ \text{smoothness term} &:= 4 \left(\sum_{i=1}^n (\alpha \gamma_{ii} + \beta_i^2) + \frac{1}{4} \sum_{i,j=1}^n (\alpha \gamma_{ij}^2 + \beta_i \beta_j \gamma_{ij} + 4\beta_i \gamma_{ij}) \right)^{1/2} \\ &\quad + \frac{1}{4} \left(\sum_{i=1}^n \beta_i^2 \right)^{1/2} \left(\sum_{i=1}^n \gamma_{ii}^2 \right)^{1/2} + 3 \sum_{i=1}^n \gamma_{ii} + \log 2, \end{aligned}$$

for

$$\begin{aligned} \alpha &:= nK + n|\log p| + n|\log(1-p)|, \\ \beta_i &:= \frac{2Kb_i}{\delta} + |\log p| + |\log(1-p)|, \quad \text{and} \\ \gamma_{ij} &:= \frac{2Kc_{ij}}{\delta} + \frac{6Kb_i b_j}{n\delta^2}. \end{aligned}$$

Moreover,

$$\log \mathbb{P}(f(Y) \geq tn) \geq -\phi_p(t + \delta_0) - \epsilon_0 n - \log 2,$$

where

$$\epsilon_0 := \frac{1}{\sqrt{n}} \left(4 + \left| \log \frac{p}{1-p} \right| \right)$$

and

$$\delta_0 := \frac{2}{n} \left(\sum_{i=1}^n (ac_{ii} + b_i^2) \right)^{1/2}.$$

We do not attempt to produce a watered down cleaner error bound, since the full power of Theorem 1.1 is needed in our applications.

1.4. Application to subgraph counts. Let $G = G(N, p)$ be an Erdős-Rényi random graph on N vertices, with edge probability p . Let H be a fixed finite simple graph. Let $\text{hom}(H, G)$ be the number of homomorphisms (edge-preserving maps) from the vertex set $V(H)$ of H into the vertex set $V(G)$ of G . This is slightly different than the number of copies of H in G , but nicer to work with mathematically. The “homomorphism density” of H in G is defined as

$$t(H, G) := \frac{\text{hom}(H, G)}{|V(G)|^{|V(H)|}}.$$

Our object of interest is the large deviation rate function for the upper tail of $t(H, G)$. Let \mathcal{P} denote upper triangular arrays like $x = (x_{ij})_{1 \leq i < j \leq N}$, where each $x_{ij} \in [0, 1]$. For any $x \in \mathcal{P}$, let G_x denote the undirected random graph whose edges are independent, and edge $\{i, j\}$ is present with probability x_{ij} , and absent with probability $1 - x_{ij}$. Let $t(H, x)$ denote the expected value of $t(H, G_x)$. Explicitly, if H has vertex set $\{1, 2, \dots, k\}$ and edge set $E(H)$, then

$$t(H, x) = \frac{1}{N^k} \sum_{q_1, \dots, q_k=1}^N \prod_{\{l, l'\} \in E(H)} x_{q_l q_{l'}}.$$

where x_{ii} is interpreted as zero for each i and $x_{ji} = x_{ij}$. For $x \in \mathcal{P}$, define

$$I_p(x) := \sum_{1 \leq i < j \leq N} I_p(x_{ij}),$$

where $I_p(x_{ij})$ is defined as in (1.3). For each $u > 1$ define

$$\psi_p(u) := \inf\{I_p(x) : x \in \mathcal{P} \text{ such that } t(H, x) \geq u \mathbb{E}(t(H, G))\}.$$

The following theorem shows that for any $u > 1$,

$$\begin{aligned} \mathbb{P}(t(H, G) \geq u \mathbb{E}(t(H, G))) \\ = \exp(-\psi_p(u) + \text{lower order terms}). \end{aligned} \tag{1.8}$$

provided that N is large and p is not too small. This approximation was proved for fixed p and N growing to infinity in [13] using Szemerédi's lemma. Various interesting consequences of this variational formula were proved in [13, 28].

Theorem 1.2. *Take any finite simple graph H and let $t(H, G)$ and ψ_p be defined as above. Let k be the number of vertices of H , m be the number of edges of H , and Δ be the maximum degree of H . Let $X := t(H, G)$. Suppose that $m \geq 1$ and $p \geq N^{-1/(m+3)}$. Then for any $u > 1$ and any N sufficiently large (depending only on H and u),*

$$1 - \frac{c(\log N)^{b_1}}{N^{b_2} p^{b_3}} \leq \frac{\psi_p(u)}{-\log \mathbb{P}(X \geq u \mathbb{E}(X))} \leq 1 + \frac{C(\log N)^{B_1}}{N^{B_2} p^{B_3}},$$

where c and C are constants that depend only on H and u , and

$$\begin{aligned} b_1 &= 1, \quad b_2 = \frac{1}{2m}, \quad b_3 = \Delta, \\ B_1 &= \frac{9+8m}{5+8m}, \quad B_2 = \frac{1}{5+8m}, \quad B_3 = \Delta - \frac{16m}{k(5+8m)}. \end{aligned}$$

For example, when H is a triangle, an explicit computation of the error terms shows that the approximation (1.8) holds whenever p goes to zero at a rate slower than $N^{-1/42}(\log N)^{11/14}$. There is no reason to believe that this should be the optimal threshold for the validity of the approximation (1.8), but at least it allows a polynomial rate of decay for p .

Shortly after the first draft of this paper was put up on arXiv, Lubetzky and Zhao [29] explicitly computed by a remarkably clever argument the limiting behavior of $\psi_p(u)$ when H is a triangle and $N^{-1} \ll p \ll 1$. With the aid of Theorem 1.2, this completely solves the large deviation problem for triangle counts when $N^{-1/42}(\log N)^{11/14} \ll p \ll 1$. Combining the solution of Lubetzky and Zhao with Theorem 1.2, we get the following corollary:

Corollary 1.3 (Corollary 1.2 in Lubetzky and Zhao [29]). *Let T be the number of triangles in $G(N, p)$. Then for any fixed $\delta > 0$,*

$$\mathbb{P}(T \geq (1 + \delta) \mathbb{E}(T)) = \exp\left(-(1 - o(1)) \min\left\{\frac{\delta^{2/3}}{2}, \frac{\delta}{3}\right\} N^2 p^2 \log \frac{1}{p}\right)$$

when $N \rightarrow \infty$ and $p \rightarrow 0$, subject to the constraint that $p \geq N^{-1/42} \log N$.

As mentioned above, [29, Theorem 1.1] gives the explicit limiting behavior of $\psi_p(u)$ whenever p goes to zero at a rate slower than N^{-1} . Therefore if one can prove a version of Theorem 1.2 that allows p to decay like $N^{-1+\epsilon}$, that would solve the problem of large deviations for triangle counts in its entirety.

The proof of Theorem 1.2 is a direct application of Theorem 1.1. The main challenge lies in verifying the low complexity gradient condition. In the case of dense graphs, the condition may be verified using Szemerédi's lemma. But it turns out that Szemerédi's lemma is not a strict requirement for proving the low complexity gradient condition for subgraph counts. One can bypass that and use a spectral argument instead. The spectral argument generalizes easily to the sparse case.

Incidentally, as already discussed in Subsection 1.1, the rough order of probability upper tails for subgraph counts drew significant interest in the random graphs community for a long time (as indicated in [20]). It was eventually determined in a series of papers by Vu [34, 35], Kim and Vu [24, 25], Janson and Ruciński [21] and finally by Janson, Oleszkiewicz and Ruciński [19]. The upper and lower bounds obtained by these authors differed by a logarithmic factor; they were matched in [10, 15] for triangle counts, and for counts of cliques in [16]. The techniques of all of these papers, however, are only suitable for getting the tail decay order and a first-order approximation such as the one given in Theorem 1.2 is not achievable by these methods.

1.5. Application to arithmetic progressions. Fixing $n \in \mathbb{N}$ and $p \in (0, 1)$, let A be a random subset of $\mathbb{Z}/n\mathbb{Z}$, constructed by keeping each element with probability p , and dropping with probability $1 - p$. In this subsection we apply Theorem 1.1 to compute large deviation probabilities for the number of three-term arithmetic progressions in A . One may be able to tackle longer arithmetic progressions via Theorem 1.1, but this would require finding a better upper bound on its complexity term.

Theorem 1.4. *Let A be a random subset of $\mathbb{Z}/n\mathbb{Z}$, constructed as above. Let X be the number of pairs $(i, j) \in (\mathbb{Z}/n\mathbb{Z})^2$ such that $\{i, i + j, i + 2j\} \subseteq A$. Let I_p be defined as in (1.4) and define*

$$\theta_p(u) := \inf \left\{ I_p(x) : x \in [0, 1]^{\mathbb{Z}/n\mathbb{Z}} \right. \\ \left. \text{such that } \sum_{i, j \in \mathbb{Z}/n\mathbb{Z}} x_i x_{i+j} x_{i+2j} \geq u \mathbb{E}(X) \right\}.$$

Suppose that $p \geq n^{-1/162}$. Then for any $u > 1$,

$$1 - cn^{-1/6} p^{-6} \log n \leq \frac{\theta_p(u)}{-\log \mathbb{P}(X \geq u \mathbb{E}(X))} \\ \leq 1 + Cn^{-1/29} p^{-162/29} (\log n)^{33/29},$$

where C and c are constants that may depend only on u .

This theorem gives an approximation for the upper tail of the number of three-term arithmetic progressions in random subsets of $\mathbb{Z}/n\mathbb{Z}$, even when the random subset is allowed to be somewhat sparse ($p \gg n^{-1/162} (\log n)^{33/162}$). Note that with $p = 1/2$, the upper tail probability is proportional to the number of subsets of $\mathbb{Z}/n\mathbb{Z}$ that contain more than a given number of three-term progressions.

Again, the main challenge in the proof of Theorem 1.4 is in establishing the low complexity gradient condition. Discrete Fourier transform techniques are used to prove that this condition holds for the function f defined above. We believe that the low complexity gradient condition should apply for longer arithmetic progressions, as it may be expected to hold in any situation where some kind of “averaging” is going on; if true, this would extend our solution to longer progressions.

The study of arithmetic progressions in subsets of integers has a long and storied history, most of which is concerned with questions of existence. An excellent survey of old and new results is

available in Tao and Vu [33]. Counting the number of sets with a given number of arithmetic progressions, or understanding the typical structure of sets that contain lots of progressions, are challenges of a different type, falling within the purview of large deviations theory. Recently a certain amount of interest has begun to grow around the resolution of such questions, quickly leading to the realization that conventional large deviations theory will not provide the answers. The most pertinent papers are the recent articles on probabilistic properties of the so-called “non-conventional averages” by Kifer [22], Kifer and Varadhan [23] and Carinci et al. [7]. In particular, Carinci et al. [7] prove a large deviation principle for what they call “two-term arithmetic progressions”, which are sums of the type $\sum x_i x_{2i}$.

1.6. Approximation of normalizing constants. Let f be as in Subsection 1.3. Consider a probability measure on $\{0, 1\}^n$ that puts mass proportional to $e^{f(x)}$ at each point x . The logarithm of the normalizing constant of this probability measure, sometimes called the “free energy”, is

$$F := \log \sum_{x \in \{0, 1\}^n} e^{f(x)}.$$

The free energy is an important object in statistical physics. In this context, the probability measure defined above is called the “Gibbs measure” with Hamiltonian f . The free energy encodes useful information about the structure of the Gibbs measure: it is often used to compute the Gibbs averages of various quantities of interest by differentiating the free energy with respect to appropriate parameters. Computation of normalizing constants is also important in statistics because it is required for computing maximum likelihood estimates of unknown parameters.

For $u \in [0, 1]$, define

$$I(u) := u \log u + (1 - u) \log(1 - u).$$

For $x = (x_1, \dots, x_n) \in [0, 1]^n$, let

$$I(x) := \sum_{i=1}^n I(x_i).$$

The goal of this subsection is to investigate conditions on f under which the approximation

$$F = \sup_{x \in [0, 1]^n} (f(x) - I(x)) + \text{lower order terms}$$

is valid. As expected from the general connection between large deviations and moment generating functions given by Varadhan’s lemma (see in [17]), the validity of the above approximation is closely related to that of (1.6).

Theorem 1.5. *Let F be defined as above, and a , b_i , c_{ij} and $\mathcal{D}(\epsilon)$ be as in Theorem 1.1. Then for any $\epsilon > 0$,*

$$F \leq \sup_{x \in [0, 1]^n} (f(x) - I(x)) + \text{complexity term} + \text{smoothness term},$$

where

$$\begin{aligned} \text{complexity term} &= \frac{1}{4} \left(n \sum_{i=1}^n b_i^2 \right)^{1/2} \epsilon + 3n\epsilon + \log |\mathcal{D}(\epsilon)|, \quad \text{and} \\ \text{smoothness term} &= 4 \left(\sum_{i=1}^n (ac_{ii} + b_i^2) + \frac{1}{4} \sum_{i,j=1}^n (ac_{ij}^2 + b_i b_j c_{ij} + 4b_i c_{ij}) \right)^{1/2} \\ &\quad + \frac{1}{4} \left(\sum_{i=1}^n b_i^2 \right)^{1/2} \left(\sum_{i=1}^n c_{ii}^2 \right)^{1/2} + 3 \sum_{i=1}^n c_{ii} + \log 2. \end{aligned}$$

Moreover, F satisfies the lower bound

$$F \geq \sup_{x \in [0,1]^n} (f(x) - I(x)) - \frac{1}{2} \sum_{i=1}^n c_{ii}.$$

Just like Theorem 1.1, it is unlikely that the error terms in Theorem 1.5 are sharp. Still, it is the first result of its kind and good enough to be applicable in some examples of interest.

Actually, Theorem 1.1 is proved in this paper as a special application Theorem 1.5. To see how this is done, take a function $g : [0,1]^n \rightarrow \mathbb{R}$ and a threshold $t \in \mathbb{R}$. Let f be a smooth function such that

$$f(x) = \begin{cases} 0 & \text{if } g(x) \geq tn, \\ \text{a large negative number} & \text{if } g(x) < tn - \text{a small quantity.} \end{cases}$$

Then $e^{f(x)}$ is a smooth approximation to the function that is 1 when $g(x) \geq tn$ and 0 when $g(x) < tn$. One may now try to apply Theorem 1.5 with this f to find an approximation to $\mathbb{P}(g(Y) \geq tn)$. This is more or less the strategy for proving Theorem 1.1 using Theorem 1.5.

The above sketch indicates that Theorem 1.5 is much more general than Theorem 1.1. Indeed, using a similar tactic it may be used for computing joint large deviations for several functions simultaneously, although we will not pursue this direction here.

1.7. Application to exponential random graphs. In this section we will use the notation of Subsection 1.4. Let l be a positive integer and H_1, \dots, H_l be finite simple graphs. Let β_1, \dots, β_l be l real numbers. Let N be another positive integer. Given a simple graph G on N vertices, let $t(H, G)$ denote, as in Subsection 1.4, the homomorphism density of H in G .

Consider the probability measure on the set of all simple graphs on N vertices that puts mass proportional to

$$\exp(N^2(\beta_1 t(H_1, G) + \dots + \beta_l t(H_l, G)))$$

on each graph G . This is an example of an exponential random graph model (ERGM). Such models are widely used in the statistical social networks community to understand the structure of networks. One of the key objectives of the practitioners is to compute estimates of the parameters β_1, \dots, β_l from an observed graph, which they assume is drawn from this model. The most popular approach to estimation is the maximum likelihood method. To implement this method, however, one needs to know the normalizing constant of the probability measure.

Until recently, the only available techniques for approximating the normalizing constants of such probability measures all relied on Markov Chain Monte Carlo (MCMC) methods. There are some doubts about the accuracy of such approximations, as pointed out in [1]. The mathematical problem was solved in [12] where it was shown that if Z_N is the normalizing constant, then as N goes to

infinity (keeping β_1, \dots, β_l are fixed),

$$\frac{\log Z_N}{N^2} \approx \sup_{x \in \mathcal{P}_N} \left(\beta_1 t(H_1, x) + \dots + \beta_l t(H_l, x) - \frac{I(x)}{N^2} \right) =: L_N,$$

where \mathcal{P}_N denotes the set \mathcal{P} defined in Subsection 1.4, that is, the set of all $x = (x_{ij})_{1 \leq i < j \leq N}$ with $x_{ij} \in [0, 1]$ for all i, j . Here the approximation sign means that the difference between the two sides tends to zero. The proof of this theorem is based on the large deviation principle for Erdős-Rényi graphs from [13]. Since this argument is based on Szemerédi's lemma, it does not give error bounds that are better than some negative power of $\log^* N$. Another problem is that this result does not allow varying the β 's with N , making it inapplicable for sparse exponential random graphs.

Theorem 1.5 solves both problems to a certain extent, by giving a concrete error bound.

Theorem 1.6. *Let Z_N and L_N be as above. Let $B := 1 + |\beta_1| + \dots + |\beta_l|$. Then*

$$\begin{aligned} -cBN^{-1} &\leq \frac{\log Z_N}{N^2} - L_N \\ &\leq CB^{8/5}N^{-1/5}(\log N)^{1/5} \left(1 + \frac{\log B}{\log N} \right) + CB^2N^{-1/2}, \end{aligned}$$

where C and c are constants that may depend only on H_1, \dots, H_l .

As an example, consider the case where $l = 2$, H_1 is a single edge, and H_2 is a triangle. In this case the above theorem shows that the difference between $N^{-2} \log Z_N$ and L_N tends to zero as long as $|\beta_1| + |\beta_2|$ grows slower than $N^{1/8}(\log N)^{-1/8}$, thereby allowing a small degree of sparsity. When the β 's are fixed, it provides an approximation error bound of order $N^{-1/5}(\log N)^{1/5}$, substantially better than the negative powers of $\log^* N$ given by Szemerédi's lemma. However, the error bound is probably suboptimal. It is an interesting challenge to figure out a sharp error bound.

2. PROOF SKETCH

In this section we give a sketch of the main ideas behind the proof of Theorem 1.5 and the main ideas behind the proof of the low complexity gradient condition for subgraph counts (which is the key ingredient in the proof of Theorem 1.2). Note that we have already sketched how Theorem 1.1 follows from Theorem 1.5 in Subsection 1.6.

We will generally denote the i th coordinate of a vector $x \in \mathbb{R}^n$ by x_i . Similarly, the i th coordinate of a random vector X will be denoted by X_i .

Let $X = (X_1, \dots, X_n)$ be a random vector that has probability density proportional to $e^{f(x)}$ on $\{0, 1\}^n$ with respect to the counting measure. For each i , define a function $\hat{x}_i : [0, 1]^n \rightarrow [0, 1]$ as

$$\hat{x}_i(x) = \mathbb{E}(X_i \mid X_j = x_j, 1 \leq j \leq n, j \neq i).$$

Let $\hat{x} : [0, 1]^n \rightarrow [0, 1]^n$ be the vector-valued function whose i th coordinate function is \hat{x}_i .

Let $\hat{X} = \hat{x}(X)$. The first step in the proof is to show that if the smoothness term in Theorem 1.5 is small, then

$$f(X) \approx f(\hat{X}) \quad \text{with high probability.} \quad (2.1)$$

(We will not bother to make precise the meaning of \approx in this sketch.) To show this, define $D := f(X) - f(\hat{X})$ and

$$h(x) := f(x) - f(\hat{x}(x)),$$

so that $D = h(X)$. For $t \in [0, 1]$ and $x \in [0, 1]^n$, let

$$u_i(t, x) := f_i(tx + (1-t)\hat{x}(x)),$$

so that

$$h(x) = \int_0^1 \sum_{i=1}^n (x_i - \hat{x}_i(x)) u_i(t, x) dt.$$

Thus,

$$\mathbb{E}(D^2) = \int_0^1 \sum_{i=1}^n \mathbb{E}((X_i - \hat{X}_i) u_i(t, X) D) dt. \quad (2.2)$$

Let $X^{(i)}$ denote the random vector $(X_1, \dots, X_{i-1}, 0, X_{i+1}, \dots, X_n)$. Let $D_i := h(X^{(i)})$. Then note that $u_i(t, X^{(i)}) D_i$ is a function of the random variables $(X_j)_{j \neq i}$ only. Therefore by the definition of \hat{X}_i ,

$$\mathbb{E}((X_i - \hat{X}_i) u_i(t, X^{(i)}) D_i) = 0.$$

Thus,

$$\begin{aligned} & \mathbb{E}((X_i - \hat{X}_i) u_i(t, X) D) \\ &= \mathbb{E}((X_i - \hat{X}_i) u_i(t, X) D) - \mathbb{E}((X_i - \hat{X}_i) u_i(t, X^{(i)}) D_i). \end{aligned}$$

If the smoothness term is small, then one can show that $u_i(t, X) \approx u_i(t, X^{(i)})$ and $D \approx D_i$. Therefore, the left-hand side of the above identity is close to zero. By (2.2), this proves the approximation (2.1).

Define a function $g : [0, 1]^n \times [0, 1]^n \rightarrow \mathbb{R}$ as

$$g(x, y) := \sum_{i=1}^n (x_i \log y_i + (1 - x_i) \log(1 - y_i)).$$

By a similar argument as above, it is possible to show that if the smoothness term is small, then

$$g(X, \hat{X}) \approx g(\hat{X}, \hat{X}) = I(\hat{X}). \quad (2.3)$$

Armed with (2.1) and (2.3), the proof of Theorem 1.5 may be completed as follows. Let A be the set of all x where $f(x) \approx f(\hat{x}(x))$ and $g(x, \hat{x}(x)) \approx I(\hat{x}(x))$. By (2.1) and (2.3), $X \in A$ with high probability. That is,

$$\frac{\sum_{x \in A} e^{f(x)}}{\sum_{x \in \{0,1\}^n} e^{f(x)}} \approx 1.$$

Therefore by the definition of the set A ,

$$\begin{aligned} F &= \log \sum_{x \in \{0,1\}^n} e^{f(x)} \approx \log \sum_{x \in A} e^{f(x)} \\ &\approx \log \sum_{x \in A} e^{f(\hat{x}(x)) - I(\hat{x}(x)) + g(x, \hat{x}(x))}. \end{aligned} \quad (2.4)$$

Now let ϵ be a small positive number, close to zero. Using the set $\mathcal{D}(\epsilon)$, it is easy to produce a set $\mathcal{D}'(\epsilon) \subseteq [0, 1]^n$ such that $|\mathcal{D}(\epsilon)| = |\mathcal{D}'(\epsilon)|$, and $\mathcal{D}'(\epsilon)$ is an ϵ -net for the image of $[0, 1]^n$ under the map \hat{x} . That is, for each x there exists $p \in \mathcal{D}'(\epsilon)$ such that

$$\sum_{i=1}^n (\hat{x}_i(x) - p_i)^2 \leq \epsilon^2 n.$$

We will say that $\hat{x}(x) \approx p$. For each $p \in \mathcal{D}'(\epsilon)$ let $\mathcal{P}(p)$ be the set of all $x \in \{0,1\}^n$ such that $\hat{x}(x) \approx p$. Then

$$\begin{aligned} & \log \sum_{x \in A} e^{f(\hat{x}(x)) - I(\hat{x}(x)) + g(x, \hat{x}(x))} \\ & \leq \log \sum_{p \in \mathcal{D}'(\epsilon)} \sum_{x \in \mathcal{P}(p)} e^{f(\hat{x}(x)) - I(\hat{x}(x)) + g(x, \hat{x}(x))} \\ & \approx \log \sum_{p \in \mathcal{D}'(\epsilon)} \sum_{x \in \mathcal{P}(p)} e^{f(p) - I(p) + g(x, p)}. \end{aligned} \tag{2.5}$$

The crucial observation is that for any $p \in [0,1]^n$,

$$\sum_{x \in \{0,1\}^n} e^{g(x,p)} = 1.$$

Thus,

$$\begin{aligned} \log \sum_{p \in \mathcal{D}'(\epsilon)} \sum_{x \in \mathcal{P}(p)} e^{f(p) - I(p) + g(x,p)} & \leq \log \sum_{p \in \mathcal{D}'(\epsilon)} e^{f(p) - I(p)} \\ & \leq \log |\mathcal{D}'(\epsilon)| + \sup_{p \in [0,1]^n} (f(p) - I(p)). \end{aligned} \tag{2.6}$$

Combining (2.4), (2.5) and (2.6) completes the proof sketch for the upper bound in Theorem 1.5.

The proof of the lower bound may be sketched as follows. Take any $y \in [0,1]^n$. Let $Y = (Y_1, \dots, Y_n)$ be a random vector with independent components, where Y_i is a *Bernoulli*(y_i) random variable. Then by Jensen's inequality,

$$\begin{aligned} \sum_{x \in \{0,1\}^n} e^{f(x)} & = \sum_{x \in \{0,1\}^n} e^{f(x) - g(x,y) + g(x,y)} \\ & = \mathbb{E}(e^{f(Y) - g(Y,y)}) \\ & \geq \exp(\mathbb{E}(f(Y) - g(Y,y))) \\ & = \exp(\mathbb{E}(f(Y)) - I(y)). \end{aligned}$$

Then, by the same line of argument that is used to prove (2.1) and (2.3), one can prove that if the error term in the lower bound is small, then $\mathbb{E}(f(Y)) \approx f(y)$. Since this is true for any y , this completes the sketch of the proof of the lower bound.

Our final task in this section is to give a sketch of the proof of the low complexity gradient condition for subgraph counts. For simplicity of exposition, let us just consider the count of triangles. Let $n = N(N-1)/2$ and let us agree to denote elements of \mathbb{R}^n as $x = (x_{ij})_{1 \leq i < j \leq N}$, with the convention that $x_{ii} = 0$ and $x_{ji} = x_{ij}$. Define a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ as

$$f(x) = \frac{1}{N} \sum_{i,j,k=1}^N x_{ij} x_{jk} x_{ki}.$$

Then note that

$$\frac{\partial f}{\partial x_{ij}} = \frac{3}{N} \sum_{k=1}^N x_{ik} x_{jk} =: 3a_{ij}(x).$$

We will now sketch why the numbers $a_{ij}(x)$ may be encoded by $o(N^2)$ bits. For any x , let $M(x)$ be the square matrix whose (i, j) th entry is x_{ij} . Note that for any x and y ,

$$\sum_{i,j=1}^N (a_{ij}(x) - a_{ij}(y))^2 = \frac{1}{N^2} \sum_{i,j,k,l} (x_{ik}x_{jk} - y_{ik}y_{jk})(x_{il}x_{jl} - y_{il}y_{jl}).$$

Let us now expand out the right-hand side and consider one pair of terms:

$$\frac{1}{N^2} \sum_{i,j,k,l} (x_{ik}x_{jk}x_{il}x_{jl} - x_{ik}x_{jk}y_{il}y_{jl}).$$

This term may be written in a telescoping manner as

$$\frac{1}{N^2} \sum_{i,j,k,l} x_{ik}x_{jk}x_{il}(x_{jl} - y_{jl}) + \frac{1}{N^2} \sum_{i,j,k,l} x_{ik}x_{jk}(x_{il} - y_{il})y_{jl}.$$

Let us consider the first term above. The crucial observation is that if i and k are fixed, then the sum in j and l is a quadratic form of the matrix $M(x) - M(y)$. Upon observing this, it is easy to see that this term is bounded above by

$$N\|M(x) - M(y)\|_{\text{op}},$$

where $\|M(x) - M(y)\|_{\text{op}}$ is the L^2 operator norm of the matrix $M(x) - M(y)$. A similar bound may be obtained for all other terms, leading to the conclusion that

$$\sum_{i,j} (a_{ij}(x) - a_{ij}(y))^2 \leq CN\|M(x) - M(y)\|_{\text{op}}, \quad (2.7)$$

where C is a universal constant.

Now take any x and let $\lambda_1, \lambda_2, \dots, \lambda_n$ be the eigenvalues of the symmetric matrix $M(x)$, arranged in decreasing order of magnitude. Then

$$\sum_{i=1}^n \lambda_i^2 = \text{Trace}(M(x)^2) = \sum_{i,j=1}^N x_{ij}^2 \leq N^2,$$

which implies the important observation that $\lambda_i^2 \leq N^2/i$ for each i since $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$. As a result of this, if M' is the matrix obtained from $M(x)$ after throwing away the terms corresponding to the $\lambda_{i+1}, \dots, \lambda_n$ in its spectral decomposition, then

$$\|M(x) - M'\|_{\text{op}} \leq \frac{N}{\sqrt{i+1}}.$$

In other words, $M(x)$ may be approximated by a rank i matrix if we allow $O(Ni^{-1/2})$ error of approximation in the operator norm. But we need only $O(Ni \log N)$ bits to encode a rank i matrix. Taking $i = \epsilon^{-4}$, and combining with the inequality (2.7), it is now easy to see how the quantities $a_{ij}(x)$ may be encoded by $O(N\epsilon^{-4} \log N)$ bits with $O(\epsilon)$ error in approximation for a typical a_{ij} , on average. This proves the low complexity gradient condition for triangle counts. The proof for general subgraph counts is a messy but straightforward generalization of the above argument.

3. PROOF OF THEOREM 1.5

In this section, we fill out the gaps in the sketch given in the previous section and thereby produce a complete proof of Theorem 1.5.

Throughout this section, we will freely use the notation of Theorem 1.5. In particular, $F, f, f_i, f_{ij}, a, b_i, c_{ij}$ and $\mathcal{D}(\epsilon)$ are as in the statement of Theorem 1.5. Let us also define some additional notation, as follows. (Some of this has already been introduced in the previous section, but we will repeat the definitions here just in case the reader has skipped that part.)

We will generally denote the i th coordinate of a vector $x \in \mathbb{R}^n$ by x_i . Similarly, the i th coordinate of a random vector X will be denoted by X_i . Given $x \in [0, 1]^n$, define $x^{(i)}$ to be the vector $(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n)$. For a random vector X define $X^{(i)}$ similarly. Given a function $g : [0, 1]^n \rightarrow \mathbb{R}$, define the discrete derivative $\Delta_i g$ as

$$\Delta_i g(x) := g(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_n) - g(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n).$$

For each i , define a function $\hat{x}_i : [0, 1]^n \rightarrow [0, 1]$ as

$$\hat{x}_i(x) = \frac{1}{1 + e^{-\Delta_i f(x)}}.$$

Let $\hat{x} : [0, 1]^n \rightarrow [0, 1]^n$ be the vector-valued function whose i th coordinate function is \hat{x}_i . When the vector x is understood from the context, we will simply write \hat{x} and \hat{x}_i instead of $\hat{x}(x)$ and $\hat{x}_i(x)$. The proof of Theorem 1.5 requires two key lemmas.

Lemma 3.1. *Let $X = (X_1, \dots, X_n)$ be a random vector that has probability density proportional to $e^{f(x)}$ on $\{0, 1\}^n$ with respect to the counting measure. Let $\hat{X} = \hat{x}(X)$. Then*

$$\mathbb{E}[(f(X) - f(\hat{X}))^2] \leq \sum_{i=1}^n (ac_{ii} + b_i^2) + \frac{1}{4} \sum_{i,j=1}^n (ac_{ij}^2 + b_i b_j c_{ij}).$$

Proof. It is easy to see that

$$\hat{x}_i(x) = \mathbb{E}(X_i \mid X_j = x_j, 1 \leq j \leq n, j \neq i).$$

Let $D := f(X) - f(\hat{X})$. Then clearly

$$|D| \leq 2a. \tag{3.1}$$

Define

$$h(x) := f(x) - f(\hat{x}(x)),$$

so that $D = h(X)$. Note that for $i \neq j$,

$$\frac{\partial \hat{x}_j}{\partial x_i} = \frac{e^{-\Delta_j f(x)}}{(1 + e^{-\Delta_j f(x)})^2} \int_0^1 f_{ij}(x_1, \dots, x_{j-1}, t, x_{j+1}, \dots, x_n) dt,$$

and for $i = j$, the above derivative is identically equal to zero. Since $e^{-x}/(1 + e^{-x})^2 \leq 1/4$ for all $x \in \mathbb{R}$, this shows that for all i and j ,

$$\left\| \frac{\partial \hat{x}_j}{\partial x_i} \right\| \leq \frac{c_{ij}}{4}. \tag{3.2}$$

Thus,

$$\begin{aligned} \left\| \frac{\partial h}{\partial x_i} \right\| &\leq \|f_i\| + \sum_{j=1}^n \|f_j\| \left\| \frac{\partial \hat{x}_j}{\partial x_i} \right\| \\ &\leq b_i + \frac{1}{4} \sum_{j=1}^n b_j c_{ij}. \end{aligned} \quad (3.3)$$

Consequently, if $D_i := h(X^{(i)})$, then

$$|D - D_i| \leq b_i + \frac{1}{4} \sum_{j=1}^n b_j c_{ij}. \quad (3.4)$$

For $t \in [0, 1]$ and $x \in [0, 1]^n$ define

$$u_i(t, x) := f_i(tx + (1-t)\hat{x}),$$

so that

$$h(x) = \int_0^1 \sum_{i=1}^n (x_i - \hat{x}_i) u_i(t, x) dt.$$

Thus,

$$\mathbb{E}(D^2) = \int_0^1 \sum_{i=1}^n \mathbb{E}((X_i - \hat{X}_i) u_i(t, X) D) dt. \quad (3.5)$$

Now,

$$\|u_i\| \leq b_i, \quad (3.6)$$

and by (3.2),

$$\begin{aligned} \left\| \frac{\partial u_i}{\partial x_i} \right\| &\leq t \|f_{ii}\| + (1-t) \sum_{j=1}^n \|f_{ij}\| \left\| \frac{\partial \hat{x}_j}{\partial x_i} \right\| \\ &\leq t c_{ii} + \frac{1-t}{4} \sum_{j=1}^n c_{ij}^2. \end{aligned} \quad (3.7)$$

The bounds (3.1), (3.4), (3.6) and (3.7) imply that

$$\begin{aligned} &|\mathbb{E}((X_i - \hat{X}_i) u_i(t, X) D) - \mathbb{E}((X_i - \hat{X}_i) u_i(t, X^{(i)}) D_i)| \\ &\leq \mathbb{E}|(u_i(t, X) - u_i(t, X^{(i)})) D| + \mathbb{E}|u_i(t, X^{(i)})(D - D_i)| \\ &\leq 2atc_{ii} + \frac{a(1-t)}{2} \sum_{j=1}^n c_{ij}^2 + b_i^2 + \frac{1}{4} \sum_{j=1}^n b_i b_j c_{ij}. \end{aligned}$$

But $u_i(t, X^{(i)}) D_i$ is a function of the random variables $(X_j)_{j \neq i}$ only. Therefore by the definition of \hat{X}_i ,

$$\mathbb{E}((X_i - \hat{X}_i) u_i(t, X^{(i)}) D_i) = 0.$$

Thus,

$$|\mathbb{E}((X_i - \hat{X}_i) u_i(t, X) D)| \leq 2atc_{ii} + \frac{a(1-t)}{2} \sum_{j=1}^n c_{ij}^2 + b_i^2 + \frac{1}{4} \sum_{j=1}^n b_i b_j c_{ij}.$$

Using this bound in (3.5) gives

$$\begin{aligned}\mathbb{E}(D^2) &\leq \int_0^1 \sum_{i=1}^n \left(2atc_{ii} + \frac{a(1-t)}{2} \sum_{j=1}^n c_{ij}^2 + b_i^2 + \frac{1}{4} \sum_{j=1}^n b_i b_j c_{ij} \right) dt \\ &= \sum_{i=1}^n (ac_{ii} + b_i^2) + \frac{1}{4} \sum_{i,j=1}^n (ac_{ij}^2 + b_i b_j c_{ij}),\end{aligned}$$

completing the proof. \square

Lemma 3.2. *Let all notation be as in Lemma 3.1. Then*

$$\mathbb{E} \left[\left(\sum_{i=1}^n (X_i - \hat{X}_i) \Delta_i f(X) \right)^2 \right] \leq \sum_{i=1}^n b_i^2 + \frac{1}{4} \sum_{i,j=1}^n b_i (b_j + 4) c_{ij}.$$

Proof. Let g_i denote the function $\Delta_i f$, for notational simplicity. Note that

$$g_i(x) = \int_0^1 f_i(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_n) dt,$$

which shows that

$$\|g_i\| \leq \|f_i\| = b_i \tag{3.8}$$

and for all j ,

$$\left\| \frac{\partial g_i}{\partial x_j} \right\| \leq \|f_{ij}\| = c_{ij}. \tag{3.9}$$

Let

$$G(x) := \sum_{i=1}^n (x_i - \hat{x}_i(x)) g_i(x).$$

Then

$$\frac{\partial G}{\partial x_i} = \sum_{j=1}^n \left[\left(1_{\{j=i\}} - \frac{\partial \hat{x}_j}{\partial x_i} \right) g_j(x) + (x_j - \hat{x}_j) \frac{\partial g_j}{\partial x_i} \right]$$

and therefore by (3.2), (3.8) and (3.9),

$$\left\| \frac{\partial G}{\partial x_i} \right\| \leq b_i + \frac{1}{4} \sum_{j=1}^n c_{ij} b_j + \sum_{j=1}^n c_{ij}. \tag{3.10}$$

Note that for any x ,

$$|G(x) - G(x^{(i)})| \leq \left\| \frac{\partial G}{\partial x_i} \right\|. \tag{3.11}$$

Again, $g_i(X)$ and $G(X^{(i)})$ are both functions of $(X_j)_{j \neq i}$ only. Therefore

$$\mathbb{E}((X_i - \hat{X}_i) g_i(X) G(X^{(i)})) = 0. \tag{3.12}$$

Combining (3.10), (3.11) and (3.12) gives

$$\begin{aligned}\mathbb{E}(G(X)^2) &= \sum_{i=1}^n \mathbb{E}((X_i - \hat{X}_i) g_i(X) G(X)) \\ &\leq \sum_{i=1}^n b_i \left(b_i + \frac{1}{4} \sum_{j=1}^n c_{ij} b_j + \sum_{j=1}^n c_{ij} \right).\end{aligned}$$

This completes the proof of the lemma. \square

With the aid of Lemma 3.1 and 3.2, we are now ready to prove Theorem 1.5.

Proof of the upper bound in Theorem 1.5. For $x, y \in [0, 1]^n$, let

$$g(x, y) := \sum_{i=1}^n (x_i \log y_i + (1 - x_i) \log(1 - y_i)).$$

Note that

$$g(x, \hat{x}) - I(\hat{x}) = \sum_{i=1}^n (x_i - \hat{x}_i) \log \frac{\hat{x}_i}{1 - \hat{x}_i} = \sum_{i=1}^n (x_i - \hat{x}_i) \Delta_i f(x). \quad (3.13)$$

Let

$$B := 4 \left(\sum_{i=1}^n (ac_{ii} + b_i^2) + \frac{1}{4} \sum_{i,j=1}^n (ac_{ij}^2 + b_i b_j c_{ij} + 4b_i c_{ij}) \right)^{1/2}.$$

Let

$$A_1 := \{x \in \{0, 1\}^n : |I(\hat{x}) - g(x, \hat{x})| \leq B/2\},$$

and

$$A_2 := \{x \in \{0, 1\}^n : |f(x) - f(\hat{x})| \leq B/2\}.$$

Let $A = A_1 \cap A_2$. By Lemma 3.1 and the identity (3.13), $\mathbb{P}(X \notin A_1) \leq 1/4$. By Lemma 3.2, $\mathbb{P}(X \notin A_2) \leq 1/4$. Thus,

$$\mathbb{P}(X \in A) \geq \frac{1}{2}.$$

That is,

$$\frac{\sum_{x \in A} e^{f(x)}}{\sum_{x \in \{0, 1\}^n} e^{f(x)}} \geq \frac{1}{2},$$

and therefore by the definition of the set A ,

$$\begin{aligned} F &= \log \sum_{x \in \{0, 1\}^n} e^{f(x)} \leq \log \sum_{x \in A} e^{f(x)} + \log 2 \\ &\leq B + \log \sum_{x \in A} e^{f(\hat{x}) - I(\hat{x}) + g(x, \hat{x})} + \log 2. \end{aligned} \quad (3.14)$$

Now take some $x \in [0, 1]^n$ and let d satisfy (1.7). Then by the Cauchy-Schwarz inequality,

$$\sum_{i=1}^n |f_i(x) - d_i| \leq n\epsilon.$$

Fix such an x and d . Note that for each i ,

$$\begin{aligned} |\Delta_i f(x) - f_i(x)| &\leq \int_0^1 |f_i(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_n) - f_i(x)| dt \\ &\leq \|f_{ii}\| = c_{ii}. \end{aligned}$$

By the last two inequalities and (1.7),

$$\sum_{i=1}^n |\Delta_i f(x) - d_i| \leq n\epsilon + \sum_{i=1}^n c_{ii}. \quad (3.15)$$

and

$$\left(\sum_{i=1}^n (\Delta_i f(x) - d_i)^2 \right)^{1/2} \leq n^{1/2} \epsilon + \left(\sum_{i=1}^n c_{ii}^2 \right)^{1/2}. \quad (3.16)$$

Let $u(x) = 1/(1 + e^{-x})$. Note that for all x ,

$$|u'(x)| = \frac{1}{(e^{x/2} + e^{-x/2})^2} \leq \frac{1}{4}.$$

Therefore if a vector $p = p(d)$ is defined as $p_i = u(d_i)$, then by (3.16),

$$\begin{aligned} \left(\sum_{i=1}^n (\hat{x}_i - p_i)^2 \right)^{1/2} &\leq \left(\frac{1}{16} \sum_{i=1}^n (\Delta_i f(x) - d_i)^2 \right)^{1/2} \\ &\leq \frac{n^{1/2} \epsilon}{4} + \frac{1}{4} \left(\sum_{i=1}^n c_{ii}^2 \right)^{1/2}. \end{aligned}$$

Thus, if

$$L := \left(\sum_{i=1}^n b_i^2 \right)^{1/2},$$

then

$$\begin{aligned} |f(\hat{x}) - f(p)| &\leq L \left(\sum_{i=1}^n (\hat{x}_i - p_i)^2 \right)^{1/2} \\ &\leq \frac{Ln^{1/2} \epsilon}{4} + \frac{L}{4} \left(\sum_{i=1}^n c_{ii}^2 \right)^{1/2}. \end{aligned} \quad (3.17)$$

Next, let $v(x) = \log(1 + e^{-x})$. Then for all x ,

$$|v'(x)| = \frac{e^{-x}}{1 + e^{-x}} \leq 1.$$

Consequently,

$$|\log \hat{x}_i - \log p_i| \leq |\Delta_i f(x) - d_i|$$

and

$$|\log(1 - \hat{x}_i) - \log(1 - p_i)| \leq |\Delta_i f(x) - d_i|.$$

Therefore by (3.15),

$$|g(x, \hat{x}) - g(x, p)| \leq 2 \sum_{i=1}^n |\Delta_i f(x) - d_i| \leq 2n\epsilon + 2 \sum_{i=1}^n c_{ii}. \quad (3.18)$$

Finally, let $w(x) = I(u(x))$. Then

$$\begin{aligned} w'(x) &= u'(x) I'(u(x)) \\ &= \frac{e^{-x}}{(1 + e^{-x})^2} \log \frac{u(x)}{1 - u(x)} \\ &= \frac{xe^{-x}}{(1 + e^{-x})^2}. \end{aligned}$$

Thus, for all x ,

$$|w'(x)| \leq \sup_{x \in \mathbb{R}} \frac{|x|e^{-x}}{(1 + e^{-x})^2} \leq \sup_{x \geq 0} xe^{-x} = \frac{1}{e}.$$

Thus,

$$|I(\hat{x}_i) - I(p_i)| \leq \frac{1}{e} |\Delta_i f(x) - d_i|,$$

and so by (3.15),

$$|I(\hat{x}) - I(p)| \leq \frac{n\epsilon}{e} + \frac{1}{e} \sum_{i=1}^n c_{ii}. \quad (3.19)$$

For each $d \in \mathcal{D}(\epsilon)$ let $\mathcal{C}(d)$ be the set of all $x \in \{0, 1\}^n$ such that (1.7) holds, and let $p(d)$ be the vector p defined above. Then by (3.17), (3.18) and (3.19),

$$\begin{aligned} \log \sum_{x \in A} e^{f(\hat{x}) - I(\hat{x}) + g(x, \hat{x})} &\leq \log \sum_{d \in \mathcal{D}(\epsilon)} \sum_{x \in \mathcal{C}(d)} e^{f(\hat{x}) - I(\hat{x}) + g(x, \hat{x})} \\ &\leq \frac{Ln^{1/2}\epsilon}{4} + \frac{L}{4} \left(\sum_{i=1}^n c_{ii}^2 \right)^{1/2} + 2n\epsilon + 2 \sum_{i=1}^n c_{ii} + \frac{n\epsilon}{e} + \frac{1}{e} \sum_{i=1}^n c_{ii} \\ &\quad + \log \sum_{d \in \mathcal{D}(\epsilon)} \sum_{x \in \mathcal{C}(d)} e^{f(p(d)) - I(p(d)) + g(x, p(d))}. \end{aligned} \quad (3.20)$$

Now note that for any $p \in [0, 1]^n$,

$$\sum_{x \in \{0, 1\}^n} e^{g(x, p)} = 1.$$

Thus,

$$\begin{aligned} \log \sum_{d \in \mathcal{D}(\epsilon)} \sum_{x \in \mathcal{C}(d)} e^{f(p(d)) - I(p(d)) + g(x, p(d))} \\ \leq \log \sum_{d \in \mathcal{D}(\epsilon)} e^{f(p(d)) - I(p(d))} \\ \leq \log |\mathcal{D}(\epsilon)| + \sup_{p \in [0, 1]^n} (f(p) - I(p)). \end{aligned} \quad (3.21)$$

Combining (3.14), (3.20) and (3.21), the proof is complete. \square

Proof of the lower bound in Theorem 1.5. Fix some $y \in [0, 1]^n$. Let $Y = (Y_1, \dots, Y_n)$ be a random vector with independent components, where Y_i is a *Bernoulli*(y_i) random variable. Then by Jensen's inequality,

$$\begin{aligned} \sum_{x \in \{0, 1\}^n} e^{f(x)} &= \sum_{x \in \{0, 1\}^n} e^{f(x) - g(x, y) + g(x, y)} \\ &= \mathbb{E}(e^{f(Y) - g(Y, y)}) \\ &\geq \exp(\mathbb{E}(f(Y) - g(Y, y))) \\ &= \exp(\mathbb{E}(f(Y)) - I(y)). \end{aligned}$$

Let $S := f(Y) - f(y)$. For $t \in [0, 1]$ and $x \in [0, 1]^n$ define

$$v_i(t, x) := f_i(tx + (1 - t)y),$$

so that

$$S = \int_0^1 \sum_{i=1}^n (Y_i - y_i) v_i(t, Y) dt. \quad (3.22)$$

By the independence of Y_i and $Y^{(i)}$,

$$\begin{aligned} |\mathbb{E}((Y_i - y_i)v_i(t, Y))| &= |\mathbb{E}((Y_i - y_i)(v_i(t, Y) - v_i(t, Y^{(i)})))| \\ &\leq \left\| \frac{\partial v_i}{\partial x_i} \right\| \leq tc_{ii}. \end{aligned}$$

Using this bound in (3.22) gives

$$\mathbb{E}(S) \geq - \int_0^1 \sum_{i=1}^n tc_{ii} dt = -\frac{1}{2} \sum_{i=1}^n c_{ii}.$$

This completes the proof. \square

4. PROOF OF THEOREM 1.1

Throughout this section, we will use the notation of Theorem 1.1 without explicit mention.

Proof of the upper bound in Theorem 1.1. Let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a function that is twice continuously differentiable, non-decreasing, and satisfies $h(x) = -1$ if $x \leq -1$ and $h(x) = 0$ if $x \geq 0$. Let $L_1 := \|h'\|$ and $L_2 := \|h''\|$. A specific choice of h is given by $h(x) = 10(x+1)^3 - 15(x+1)^4 + 6(x+1)^5 - 1$ for $-1 \leq x \leq 0$, which gives $L_1 \leq 2$ and $L_2 \leq 6$. Define

$$\psi(x) := Kh((x-t)/\delta).$$

Then clearly

$$\|\psi\| \leq K, \quad \|\psi'\| \leq \frac{L_1 K}{\delta}, \quad \|\psi''\| \leq \frac{L_2 K}{\delta^2}.$$

Let

$$g(x) := n\psi(f(x)/n) + \sum_{i=1}^n (x_i \log p + (1-x_i) \log(1-p)).$$

The plan is to apply Theorem 1.5 to the function g instead of f . Note that $\psi(x) = 0$ if $x \geq t$. Thus,

$$\begin{aligned} \mathbb{P}(f(Y) \geq tn) &\leq \mathbb{E}(e^{n\psi(f(Y)/n)}) \\ &= \sum_{x \in \{0,1\}^n} e^{g(x)}. \end{aligned}$$

Note also that for any $x \in [0, 1]^n$ such that $f(x) \geq tn$,

$$g(x) - I(x) = n\psi(f(x)/n) - I_p(x) = -I_p(x) \leq -\phi_p(t).$$

Again, if $f(x) \leq (t - \delta)n$, then $(f(x)/n - t)/\delta \leq -1$, and so

$$g(x) - I(x) = -nK - I_p(x) \leq -nK = -\phi_p(t).$$

Finally, note that if $f(x) = (t - \delta')n$ for some $0 < \delta' < \delta$, then

$$g(x) - I(x) \leq -I_p(x) \leq -\phi_p(t - \delta') \leq -\phi(t - \delta).$$

Thus,

$$\sup_x (g(x) - I(x)) \leq -\phi_p(t - \delta).$$

Let $C_p := |\log p| + |\log(1-p)|$. Note that

$$\|g\| \leq nK + nC_p = \alpha,$$

and for any i ,

$$\left\| \frac{\partial g}{\partial x_i} \right\| \leq \frac{2Kb_i}{\delta} + C_p = \beta_i,$$

and for any i, j ,

$$\left\| \frac{\partial^2 g}{\partial x_i \partial x_j} \right\| \leq \frac{2Kc_{ij}}{\delta} + \frac{6Kb_i b_j}{n\delta^2} = \gamma_{ij}.$$

Next, fix some $\epsilon > 0$ and let $\mathcal{D}(\epsilon)$ be as in Section 3. Let

$$\epsilon' := \frac{\epsilon}{2\|\psi'\|}, \quad \tau := \frac{\epsilon}{2\left(\frac{1}{n} \sum_{i=1}^n b_i^2\right)^{1/2}}.$$

Let $l \in \mathbb{R}^n$ be the vector whose coordinates are all equal to $\log(p/(1-p))$ and define

$$\mathcal{D}'(\epsilon) := \{\theta d + l : d \in \mathcal{D}(\epsilon'), \theta = j\tau \text{ for some integer } 0 \leq j < \|\psi'\|/\tau\}.$$

Let $g_i := \partial g / \partial x_i$. Take any $x \in [0, 1]^n$, and choose $d \in \mathcal{D}(\epsilon)$ satisfying (1.7). Choose an integer j between 0 and $\|\psi'\|/\tau$ such that $|\psi'(f(x)/n) - j\tau| \leq \tau$. Let $d' := j\tau d + l$, so that $d' \in \mathcal{D}'(\epsilon)$. Then

$$\begin{aligned} \sum_{i=1}^n (g_i(x) - d'_i)^2 &= \sum_{i=1}^n (\psi'(f(x)/n) f_i(x) - j\tau d_i)^2 \\ &\leq 2(\psi'(f(x)/n) - j\tau)^2 \sum_{i=1}^n f_i(x)^2 + 2\|\psi'\|^2 \sum_{i=1}^n (f_i(x) - d_i)^2 \\ &\leq 2\tau^2 \sum_{i=1}^n b_i^2 + 2\|\psi'\|^2 n\epsilon'^2 = n\epsilon^2. \end{aligned}$$

This shows that $\mathcal{D}'(\epsilon)$ plays the role of $\mathcal{D}(\epsilon)$ for the function g . Note that

$$|\mathcal{D}'(\epsilon)| \leq \frac{\|\psi'\|}{\tau} |\mathcal{D}(\epsilon')|.$$

This gives the upper bound on the complexity term for the function g . The proof is completed by applying Theorem 1.5. \square

Proof of the lower bound in Theorem 1.1. Fix any $z \in [0, 1]^n$ such that

$$f(z) \geq (t + \delta_0)n.$$

Let $Z = (Z_1, \dots, Z_n)$ be a random vector with independent components, where $Z_i \sim \text{Bernoulli}(z_i)$. Let \mathcal{A} be the set of all $x \in \{0, 1\}^n$ such that $f(x) \geq tn$. Let \mathcal{A}' be the subset of \mathcal{A} where $|g(x, z) - g(x, p) - I_p(z)| \leq \epsilon_0 n$. Then

$$\begin{aligned} \mathbb{P}(f(Y) \geq tn) &= \sum_{x \in \mathcal{A}} e^{g(x, p)} \\ &= \sum_{x \in \mathcal{A}} e^{g(x, p) - g(x, z) + g(x, z)} \\ &\geq \sum_{x \in \mathcal{A}'} e^{g(x, p) - g(x, z) + g(x, z)} \geq e^{-I_p(z) - \epsilon_0 n} \mathbb{P}(Z \in \mathcal{A}'). \end{aligned} \tag{4.1}$$

Note that

$$\mathbb{E}(g(Z, z) - g(Z, p)) = I_p(z),$$

and

$$\begin{aligned}
& \text{Var}(g(Z, z) - g(Z, p)) \\
&= \sum_{i=1}^n \text{Var}(Z_i \log(z_i/p) + (1 - Z_i) \log((1 - z_i)/(1 - p))) \\
&= \sum_{i=1}^n z_i(1 - z_i) \left(\log \frac{z_i/p}{(1 - z_i)/(1 - p)} \right)^2.
\end{aligned}$$

Using the inequalities $|\sqrt{x} \log x| \leq 2/e \leq 1$ and $x(1 - x) \leq 1/4$, we see that for any $x \in [0, 1]$,

$$\begin{aligned}
& x(1 - x) \left(\log \frac{x/p}{(1 - x)/(1 - p)} \right)^2 \\
& \leq \left(|\sqrt{x} \log x| + |\sqrt{1 - x} \log(1 - x)| + \frac{1}{2} \left| \log \frac{p}{1 - p} \right| \right)^2 \\
& \leq \left(2 + \frac{1}{2} \left| \log \frac{p}{1 - p} \right| \right)^2.
\end{aligned}$$

Combining the last three displays, we see that

$$\mathbb{P}(|g(Z, z) - g(Z, p) - I_p(z)| > \epsilon_0 n) \leq \frac{1}{\epsilon_0^2 n} \left(2 + \frac{1}{2} \left| \log \frac{p}{1 - p} \right| \right)^2 = \frac{1}{4}. \quad (4.2)$$

Let $S := f(Z) - f(z)$ and $v_i(t, x) := f_i(tZ + (1 - t)z)$. Let $S_i := f(Z^{(i)}) - f(z)$, so that $|S - S_i| \leq b_i$. Since

$$S = \int_0^1 \sum_{i=1}^n (Z_i - z_i) v_i(t, Z) dt,$$

we have

$$\mathbb{E}(S^2) = \int_0^1 \sum_{i=1}^n \mathbb{E}((Z_i - z_i) v_i(t, Z) S) dt. \quad (4.3)$$

By the independence of Z_i and the pair $(S_i, Z^{(i)})$,

$$\begin{aligned}
& |\mathbb{E}((Z_i - z_i) v_i(t, Z) S)| \\
&= |\mathbb{E}((Z_i - z_i) (v_i(t, Z) S - v_i(t, Z^{(i)}) S_i))| \\
&\leq \|S\| \left\| \frac{\partial v_i}{\partial x_i} \right\| + \|v_i\| \|S - S_i\| \\
&\leq 2atc_{ii} + b_i^2.
\end{aligned}$$

By (4.3), this gives

$$\mathbb{E}(S^2) \leq \sum_{i=1}^n (ac_{ii} + b_i^2).$$

Therefore,

$$\mathbb{P}(f(Z) < tn) \leq \frac{1}{\delta_0^2 n^2} \sum_{i=1}^n (ac_{ii} + b_i^2) = \frac{1}{4}. \quad (4.4)$$

Inequalities (4.2) and (4.4) give

$$\mathbb{P}(Z \in \mathcal{A}') \geq \frac{1}{2}.$$

Plugging this into (4.1) and taking supremum over z completes the proof. \square

5. PROOF OF THEOREM 1.2

Let all notation be the same as in the statement of Theorem 1.2. Let

$$n := \binom{N}{2}.$$

Throughout this section, we will index the elements of \mathbb{R}^n as

$$x = (x_{ij})_{1 \leq i < j \leq N},$$

with the understanding that if $i < j$, then x_{ji} is the same as x_{ij} , and for all i , $x_{ii} = 0$. Let k be a positive integer, and let H be a finite simple graph on the vertex set $[k] := \{1, \dots, k\}$. Let E be the set of edges of H and let $m := |E|$.

Define a function $T : [0, 1]^n \rightarrow \mathbb{R}$ as

$$T(x) := \frac{1}{N^{k-2}} \sum_{q \in [N]^k} \prod_{\{l, l'\} \in E} x_{qlq_{l'}}, \quad (5.1)$$

so that $t(H, G_x) = T(x)/N^2$. The plan is to apply Theorem 1.1 with $f = T$. We will now compute the required bounds for the function T .

Lemma 5.1. *For the function T on \mathbb{R}^n defined above, $\|T\| \leq N^2$, and for any $i < j$ and $i' < j'$,*

$$\begin{aligned} \left\| \frac{\partial T}{\partial x_{ij}} \right\| &\leq 2m, \text{ and} \\ \left\| \frac{\partial^2 T}{\partial x_{ij} \partial x_{i'j'}} \right\| &\leq \begin{cases} 4m(m-1)N^{-1} & \text{if } |\{i, j, i', j'\}| = 2 \text{ or } 3, \\ 4m(m-1)N^{-2} & \text{if } |\{i, j, i', j'\}| = 4. \end{cases} \end{aligned}$$

Proof. It is clear that $\|T\| \leq N^2$ since the x_{ij} 's are all in $[0, 1]$ and there are exactly N^k terms in the sum that defines T . Next, note that for any $i < j$,

$$\frac{\partial T}{\partial x_{ij}} = \frac{1}{N^{k-2}} \sum_{\{a, b\} \in E} \sum_{\substack{q \in [N]^k \\ \{q_a, q_b\} = \{i, j\}}} \prod_{\substack{\{l, l'\} \in E \\ \{l, l'\} \neq \{a, b\}}} x_{qlq_{l'}}, \quad (5.2)$$

and therefore

$$\left\| \frac{\partial T}{\partial x_{ij}} \right\| \leq \frac{2mN^{k-2}}{N^{k-2}} = 2m.$$

Next, for any $i < j$ and $i' < j'$,

$$\frac{\partial^2 T}{\partial x_{ij} \partial x_{i'j'}} = \frac{1}{N^{k-2}} \sum_{\{a, b\} \in E} \sum_{\substack{\{c, d\} \in E \\ \{c, d\} \neq \{a, b\}}} \sum_{\substack{q \in [N]^k \\ \{q_a, q_b\} = \{i, j\} \\ \{q_c, q_d\} = \{i', j'\}}} \prod_{\substack{\{l, l'\} \in E \\ \{l, l'\} \neq \{a, b\} \\ \{l, l'\} \neq \{c, d\}}} x_{qlq_{l'}}.$$

Take any two edges $\{a, b\}, \{c, d\} \in E$ such that $\{a, b\} \neq \{c, d\}$. Then the number of choices of $q \in [N]^k$ such that $\{q_a, q_b\} = \{i, j\}$ and $\{q_c, q_d\} = \{i', j'\}$ is at most $4N^{k-3}$ if $|\{i, j, i', j'\}| = 2$ or 3 (since we are constraining q_a, q_b, q_c and q_d and $|\{a, b, c, d\}| \geq 3$ always), and at most $4N^{k-4}$ if

$|\{i, j, i', j'\}| = 4$ (since $|\{a, b, c, d\}|$ must be 4 if there is at least one possible choice of q for these i, j, i', j'). This gives the upper bound for the second derivatives. \square

Lemma 5.2. *For the function T defined above, one can produce sets $\mathcal{D}(\epsilon)$ satisfying the criterion (1.7) (with $f = T$) such that*

$$|\mathcal{D}(\epsilon)| \leq \exp\left(\frac{C_1 m^4 k^4 N}{\epsilon^4} \log \frac{C_2 m^4 k^4}{\epsilon^4}\right),$$

where C_1 and C_2 are universal constants.

The proof of Lemma 5.2 requires some preparation. We begin by introducing some special notation. For an $N \times N$ matrix M , recall the definition of the operator norm:

$$\|M\|_{\text{op}} := \max\{\|Mx\| : x \in \mathbb{R}^N, \|x\| = 1\}.$$

For $x = (x_{ij})_{1 \leq i < j \leq N} \in \mathbb{R}^n$, let $M(x)$ be the symmetric matrix whose (i, j) th entry is x_{ij} , with the convention that $x_{ij} = x_{ji}$ and $x_{ii} = 0$. Define the operator norm on \mathbb{R}^n as

$$\|x\|_{\text{op}} := \|M(x)\|_{\text{op}}.$$

The following lemma estimates the entropy of the unit cube under this norm.

Lemma 5.3. *For any $\tau \in (0, 1)$, there is a finite set of $N \times N$ matrices $\mathcal{W}(\tau)$ such that*

$$|\mathcal{W}(\tau)| \leq e^{34(N/\tau^2) \log(51/\tau^2)},$$

and for any $N \times N$ matrix M with entries in $[0, 1]$, there exists $W \in \mathcal{W}(\tau)$ such that

$$\|M - W\|_{\text{op}} \leq N\tau.$$

In particular, for any $x \in [0, 1]^n$ there exists $W \in \mathcal{W}(\tau)$ such that $\|M(x) - W\|_{\text{op}} \leq N\tau$.

Proof. Let l be the integer part of $17/\tau^2$ and $\delta = 1/l$. Let \mathcal{A} be a finite subset of the unit ball of \mathbb{R}^N such that any vector inside the ball is at Euclidean distance $\leq \delta$ from some element of \mathcal{A} . (In other words, \mathcal{A} is a δ -net of the unit ball under the Euclidean metric.) The set \mathcal{A} may be defined as a maximal set of points in the unit ball such that any two are at a distance greater than δ from each other. Since the balls of radius $\delta/2$ around these points are disjoint and their union is contained in the ball of radius $1 + \delta/2$ centered at zero, it follows that $|\mathcal{A}|C(\delta/2)^N \leq C(1 + \delta/2)^N$, where C is the volume of the unit ball. Therefore,

$$|\mathcal{A}| \leq (3/\delta)^N. \tag{5.3}$$

Take any $x \in \mathbb{R}^n$. Suppose that M has singular value decomposition

$$M = \sum_{i=1}^n \lambda_i u_i v_i^t,$$

where $\lambda_1 \geq \lambda_2 \geq \dots \lambda_n \geq 0$ are the singular values of M , and u_1, \dots, u_n and v_1, \dots, v_n are singular vectors, and v_i^t denotes the transpose of the column vector v_i . Assume that the u_i 's and v_i 's are orthonormal systems. Since the elements of M all belong to the interval $[0, 1]$, it is easy to see that $\lambda_1 \leq N$ and $\sum \lambda_i^2 \leq N^2$. Due to the second inequality, there exists $y \in \mathcal{A}$ such that

$$\sum_{i=1}^N (N^{-1} \lambda_i - y_i)^2 \leq \delta^2. \tag{5.4}$$

Let z_1, \dots, z_N and w_1, \dots, w_N be elements of \mathcal{A} such that for each i ,

$$\sum_{j=1}^N (u_{ij} - z_{ij})^2 \leq \delta^2 \quad \text{and} \quad \sum_{j=1}^N (v_{ij} - w_{ij})^2 \leq \delta^2, \quad (5.5)$$

where u_{ij} denotes the j th component of the vector u_i , etc. Define two matrices V and W as

$$V := \sum_{i=1}^{l-1} \lambda_i u_i v_i^t \quad \text{and} \quad W := \sum_{i=1}^{l-1} N y_i z_i w_i^t.$$

Note that since $\sum \lambda_i^2 \leq N^2$ and λ_i decreases with i , therefore for each i , $\lambda_i^2 \leq N^2/i$. Thus,

$$\begin{aligned} \|M - W\|_{\text{op}} &\leq \|M - V\|_{\text{op}} + \|V - W\|_{\text{op}} \\ &\leq \frac{N}{\sqrt{l}} + \|V - W\|_{\text{op}}. \end{aligned}$$

Next, note that by (5.5), the operator norms of the rank-one matrices $(u_i - z_i)v_i^t$ and $z_i(v_i - w_i)^t$ are bounded by δ . And by (5.4), $|\lambda_i - N y_i| \leq N\delta$ for each i . Therefore

$$\begin{aligned} \|V - W\|_{\text{op}} &\leq \left\| \sum_{i=1}^{l-1} (\lambda_i - N y_i) u_i v_i^t \right\|_{\text{op}} + \left\| \sum_{i=1}^{l-1} N y_i (u_i - z_i) v_i^t \right\|_{\text{op}} \\ &\quad + \left\| \sum_{i=1}^{l-1} N y_i z_i (v_i - w_i)^t \right\|_{\text{op}} \\ &\leq \max_{1 \leq i \leq l-1} |\lambda_i - N y_i| + 2 \sum_{i=1}^{l-1} N |y_i| \delta \\ &\leq N\delta + 2N\delta \left((l-1) \sum_{i=1}^{l-1} y_i^2 \right)^{1/2} \leq N\delta + 2N\delta \sqrt{l-1} \leq \frac{3N}{\sqrt{l}}. \end{aligned}$$

Thus,

$$\|M - W\|_{\text{op}} \leq \frac{4N}{\sqrt{l}} \leq \frac{4N}{\sqrt{\frac{17}{\tau^2} - 1}} \leq \frac{4N}{\sqrt{\frac{16}{\tau^2}}} = N\tau.$$

Let $\mathcal{W}(\tau)$ be the set of all possible W 's constructed in the above manner. Then $\mathcal{W}(\tau)$ has the required property, and by (5.3),

$$\begin{aligned} |\mathcal{W}(\tau)| &\leq \text{The number of ways of choosing} \\ &\quad y, z_1, \dots, z_{l-1}, w_1, \dots, w_{l-1} \in \mathcal{A} \\ &= |\mathcal{A}|^{2l-1} \leq (3/\delta)^{2Nl} = e^{2Nl \log(3l)}. \end{aligned}$$

This completes the proof of the lemma. \square

Let r be a positive integer. Let K_r be the complete graph on the vertex set $\{1, \dots, r\}$. For any set of edges A of K_r , any $q = (q_1, \dots, q_r) \in [N]^r$, and any $x \in [0, 1]^n$, let

$$P(x, q, A) := \prod_{\{a, b\} \in A} x_{q_a q_b},$$

with the usual convention that the empty product is 1. Note that if $q_a = q_b$ for some $\{a, b\} \in A$, the $P(x, q, A) = 0$ due to our convention that $x_{ii} = 0$ for each i . Next, note that if A and B are disjoint sets of edges, then

$$P(x, q, A \cup B) = P(x, q, A)P(x, q, B). \quad (5.6)$$

Lemma 5.4. *Let A and B be sets of edges of K_r , and let $e = \{\alpha, \beta\}$ be an edge that is neither in A nor in B . Then for any $x, y \in [0, 1]^n$,*

$$\left| \sum_{q \in [N]^r} P(x, q, A)P(y, q, B)(x_{q_\alpha q_\beta} - y_{q_\alpha q_\beta}) \right| \leq N^{r-1} \|x - y\|_{\text{op}}.$$

Proof. By relabeling the vertices of K_r and redefining A and B , we may assume that $\alpha = 1$ and $\beta = 2$.

Let A_1 be the set of all edges in A that are incident to 1. Let A_2 be the set of all edges in A that are incident to 2. Note that since $\{1, 2\} \notin A$, therefore A_1 and A_2 must be disjoint. Similarly, let B_1 be the set of all edges in B that are incident to 1 and let B_2 be the set of all edges in B that are incident to 2. Let $A_3 = A \setminus (A_1 \cup A_2)$ and $B_3 = B \setminus (B_1 \cup B_2)$. By (5.6),

$$P(x, q, A) = P(x, q, A_1)P(x, q, A_2)P(x, q, A_3)$$

and

$$P(y, q, B) = P(y, q, B_1)P(y, q, B_2)P(y, q, B_3).$$

Thus,

$$\begin{aligned} & \sum_{q \in [N]^r} P(x, q, A)P(y, q, B)(x_{q_1 q_2} - y_{q_1 q_2}) \\ &= \sum_{q_3, \dots, q_r} P(x, q, A_3)P(y, q, B_3) \left(\sum_{q_1, q_2} Q(x, y, q)(x_{q_1 q_2} - y_{q_1 q_2}) \right), \end{aligned}$$

where

$$Q(x, y, q) = P(x, q, A_1)P(x, q, A_2)P(y, q, B_1)P(y, q, B_2).$$

Now fix q_3, \dots, q_r . Then $P(x, q, A_1)P(y, q, B_1)$ is a function of q_1 only, and does not depend on q_2 . Let $g(q_1)$ denote this function. Similarly, $P(x, q, A_2)P(y, q, B_2)$ is a function of q_2 only, and does not depend on q_1 . Let $h(q_2)$ denote this function. Both g and h are uniformly bounded by 1. Therefore

$$\begin{aligned} \left| \sum_{q_1, q_2} Q(x, y, q)(x_{q_1 q_2} - y_{q_1 q_2}) \right| &= \left| \sum_{q_1, q_2} g(q_1)h(q_2)(x_{q_1 q_2} - y_{q_1 q_2}) \right| \\ &\leq N \|x - y\|_{\text{op}}. \end{aligned}$$

Since this is true for all choices of q_3, \dots, q_r and P is also uniformly bounded by 1, this completes the proof of the lemma. \square

Let A and B be two sets of edges of K_r . For $x, y \in [0, 1]^n$, define

$$R(x, y, A, B) := \sum_{q \in [N]^r} P(x, q, A)P(y, q, B).$$

Lemma 5.5. *Let A, B, A' and B' be sets of edges of K_r such that $A \cap B = A' \cap B' = \emptyset$ and $A \cup B = A' \cup B'$. Then*

$$|R(x, y, A, B) - R(x, y, A', B')| \leq \frac{1}{2} r(r-1) N^{r-1} \|x - y\|_{\text{op}}.$$

Proof. First, suppose that $e = \{\alpha, \beta\}$ is an edge such that $e \notin A'$ and $A = A' \cup \{e\}$. Since $A \cup B = A' \cup B'$ and $A \cap B = A' \cap B' = \emptyset$, this implies that $e \notin B$ and $B' = B \cup \{e\}$. Thus,

$$R(x, y, A, B) - R(x, y, A', B') = \sum_{q \in [N]^r} P(x, q, A') P(y, q, B) (x_{q_\alpha q_\beta} - y_{q_\alpha q_\beta}),$$

and the proof is completed using Lemma 5.4. For the general case, simply ‘move’ from the pair (A, B) to the pair (A', B') by ‘moving one edge at a time’ and apply Lemma 5.4 at each step. \square

Lemma 5.6. *Let g_{ij} denote the function $\partial T / \partial x_{ij}$, where T is the function defined in equation (5.1). Then for any $x, y \in [0, 1]^n$,*

$$\sum_{1 \leq i < j \leq N} (g_{ij}(x) - g_{ij}(y))^2 \leq 8m^2 k^2 N \|x - y\|_{\text{op}}.$$

Proof. Recall equation (5.2), that is, for any $1 \leq i < j \leq N$,

$$g_{ij}(x) = \frac{\partial T}{\partial x_{ij}} = \frac{1}{N^{k-2}} \sum_{\{a, b\} \in E} \sum_{\substack{q \in [N]^k \\ \{q_a, q_b\} = \{i, j\}}} \prod_{\substack{\{l, l'\} \in E \\ \{l, l'\} \neq \{a, b\}}} x_{q_l q_{l'}}.$$

Although differentiating with respect to x_{ii} does not make sense, let g_{ii} be the function defined using the same formula as above. When $i > j$, let $g_{ij} = g_{ji}$. Fix $x, y \in [0, 1]^n$. Define for any $q \in [N]^k$ and $\{a, b\} \in E$

$$D(q, \{a, b\}) := \prod_{\substack{\{l, l'\} \in E \\ \{l, l'\} \neq \{a, b\}}} x_{q_l q_{l'}} - \prod_{\substack{\{l, l'\} \in E \\ \{l, l'\} \neq \{a, b\}}} y_{q_l q_{l'}}.$$

Define

$$\theta_{ij} := \begin{cases} 2 & \text{if } i = j, \\ 1/2 & \text{if } i \neq j, \end{cases} \quad \gamma_{ij} := \begin{cases} 2 & \text{if } i = j, \\ 1 & \text{if } i \neq j. \end{cases}$$

Then note that

$$\begin{aligned} & \sum_{i, j=1}^N \theta_{ij} (g_{ij}(x) - g_{ij}(y))^2 \\ &= \frac{1}{N^{2k-4}} \sum_{i, j=1}^N \theta_{ij} \left(\sum_{\{a, b\} \in E} \sum_{\substack{q \in [N]^k \\ \{q_a, q_b\} = \{i, j\}}} D(q, \{a, b\}) \right)^2 \\ &= \frac{1}{N^{2k-4}} \sum_{i, j=1}^N \sum_{\substack{\{a, b\} \in E \\ \{c, d\} \in E}} \sum_{\substack{q \in [N]^k \\ \{q_a, q_b\} = \{i, j\}}} \sum_{\substack{s \in [N]^k \\ \{s_c, s_d\} = \{i, j\}}} \theta_{ij} D(q, \{a, b\}) D(s, \{c, d\}) \\ &= \frac{1}{N^{2k-4}} \sum_{\substack{\{a, b\} \in E \\ \{c, d\} \in E}} \sum_{q \in [N]^k} \sum_{\substack{s \in [N]^k \\ \{s_c, s_d\} = \{q_a, q_b\}}} \gamma_{q_a q_b} D(q, \{a, b\}) D(s, \{c, d\}). \end{aligned}$$

Now fix two edges $\{a, b\}$ and $\{c, d\}$ in E . Relabeling vertices if necessary, assume that $c = k - 1$ and $d = k$. Let $r = 2k - 2$. For any $t \in [N]^r$, define two vectors $q(t)$ and $s(t)$ in $[N]^k$ as follows.

For $i = 1, \dots, k$, let $q_i(t) = t_i$. For $i = 1, \dots, k-2$, let $s_i(t) = t_{i+k}$. Let $s_{k-1}(t) = t_a$ and $s_k(t) = t_b$. With this definition, it is clear that

$$\begin{aligned} & \sum_{q \in [N]^k} \sum_{\substack{s \in [N]^k \\ \{s_c, s_d\} = \{q_a, q_b\}}} \gamma_{q_a q_b} D(q, \{a, b\}) D(s, \{c, d\}) \\ &= \sum_{q \in [N]^k} \sum_{\substack{s \in [N]^k \\ s_c = q_a, s_d = q_b}} D(q, \{a, b\}) D(s, \{c, d\}) \\ &+ \sum_{q \in [N]^k} \sum_{\substack{s \in [N]^k \\ s_c = q_b, s_d = q_a}} D(q, \{a, b\}) D(s, \{c, d\}). \end{aligned}$$

Note that the first term on the right-hand side is exactly equal to

$$\sum_{t \in [N]^r} D(q(t), \{a, b\}) D(s(t), \{c, d\}).$$

Below, we will get a bound on this term. The same upper bound will hold for the other term by symmetry.

Next, define two subsets of edges A and B of K_r as follows. Let A be the set of all edges $\{l, l'\}$ such that $\{l, l'\} \in E \setminus \{\{a, b\}\}$. Let B be the set of all edges $\{\phi(l), \phi(l')\}$ such that $\{l, l'\} \in E \setminus \{\{k-1, k\}\}$, where $\phi : [k] \rightarrow [r]$ is the map

$$\phi(x) = \begin{cases} x+k & \text{if } x \neq k-1 \text{ and } x \neq k, \\ a & \text{if } x = k-1, \\ b & \text{if } x = k. \end{cases}$$

By the above construction, $q_l(t) = t_l$ and $s_l(t) = t_{\phi(l)}$. Therefore it is easy to see, for instance, that

$$\begin{aligned} & \sum_{t \in [N]^r} \prod_{\substack{\{l, l'\} \in E \\ \{l, l'\} \neq \{a, b\}}} x_{q_l(t) q_{l'}(t)} \prod_{\substack{\{l, l'\} \in E \\ \{l, l'\} \neq \{k-1, k\}}} y_{s_l(t) s_{l'}(t)} \\ &= \sum_{t \in [N]^r} \prod_{\substack{\{l, l'\} \in E \\ \{l, l'\} \neq \{a, b\}}} x_{t_l t_{l'}} \prod_{\substack{\{l, l'\} \in E \\ \{l, l'\} \neq \{k-1, k\}}} y_{t_{\phi(l)} t_{\phi(l')}} \\ &= R(x, y, A, B). \end{aligned}$$

Carrying out similar computations for the other terms in the expansion of $D(q(t), \{a, b\}) D(s(t), \{c, d\})$, we get

$$\begin{aligned} & \sum_{t \in [N]^r} D(q(t), \{a, b\}) D(s(t), \{c, d\}) \\ &= R(x, y, A \cup B, \emptyset) - R(x, y, B, A) - R(x, y, A, B) + R(x, y, \emptyset, A \cup B). \end{aligned}$$

Lastly, note that $A \cap B = \emptyset$ since for any $\{l, l'\} \in E \setminus \{\{k-1, k\}\}$, at least one among $\phi(l)$ and $\phi(l')$ must be strictly bigger than k and therefore $\{\phi(l), \phi(l')\}$ cannot be an element of A . The proof is now easily completed by applying Lemma 5.5. \square

With the help of Lemma 5.3 and Lemma 5.6, we are now ready to prove Lemma 5.2.

Proof of Lemma 5.2. Take any $\epsilon > 0$ and let

$$\tau = \frac{\epsilon^2}{64m^2k^2}.$$

Let $\mathcal{W}(\tau)$ be as in Lemma 5.3. For each $W \in \mathcal{W}(\tau)$, let $y(W) \in [0, 1]^n$ be a vector such that $\|M(y) - W\|_{\text{op}} \leq N\tau$. If for some W there does not exist any such y , leave $y(W)$ undefined. Let $g_{ij} = \partial T / \partial x_{ij}$, as in Lemma 5.6. Let $g : [0, 1]^n \rightarrow \mathbb{R}^n$ be the function whose (i, j) th coordinate is g_{ij} . Define

$$\mathcal{D}(\epsilon) := \{g(y) : y = y(W) \text{ for some } W \in \mathcal{W}(\tau)\}.$$

Then by Lemma 5.3

$$|\mathcal{D}(\epsilon)| \leq |\mathcal{W}(\tau)| \leq e^{34(N/\tau^2) \log(51/\tau^2)}.$$

We claim that the set $\mathcal{D}(\epsilon)$ satisfies the requirements of Theorem 1.1. To see this, take any $x \in [0, 1]^n$. By Lemma 5.3, there exists $W \in \mathcal{W}(\tau)$ such that $\|M(x) - W\|_{\text{op}} \leq N\tau$. In particular, this means that $y := y(W)$ is defined, and so

$$\begin{aligned} \|x - y\|_{\text{op}} &= \|M(x) - M(y)\|_{\text{op}} \\ &\leq \|M(x) - W\|_{\text{op}} + \|W - M(y)\|_{\text{op}} \\ &\leq 2N\tau. \end{aligned}$$

Therefore by Lemma 5.6,

$$\sum_{1 \leq i < j \leq N} (g_{ij}(x) - g_{ij}(y))^2 \leq 16m^2k^2N^2\tau.$$

Let $z = g(x)$ and $v = g(y)$. Then $v \in \mathcal{D}(\epsilon)$, and by the above inequality,

$$\sum_{1 \leq i < j \leq N} (z_{ij} - v_{ij})^2 \leq 16m^2k^2N^2\tau = \frac{N^2\epsilon^2}{4} \leq \binom{N}{2}\epsilon^2.$$

This proves the claim that $\mathcal{D}(\epsilon)$ satisfies the requirements of Theorem 1.1. This completes the proof of Lemma 5.2. \square

The next step is to understand the properties of the rate function $\phi_p(t)$ corresponding to T .

Lemma 5.7. *Let $\phi_p(t)$ be defined as in (1.5), with $f = T$ and $n = N(N-1)/2$. Let l be the element of $[0, 1]^n$ whose coordinates are all equal to 1, and let $t_0 := T(l)/n$. Then for any $0 < \delta < t < t_0$,*

$$\phi_p(t - \delta) \geq \phi_p(t) - \left(\frac{\delta}{t_0 - t} \right)^{1/m} n \log(1/p).$$

Proof. Take any $x \in [0, 1]^n$ such that $T(x) \geq (t - \delta)n$ and x minimizes $I_p(x)$ among all x satisfying this inequality. If $T(x) \geq tn$, then we immediately have $\phi_p(t) \leq I_p(x) = \phi_p(t - \delta)$, and there is nothing more to prove. So let us assume that $T(x) < tn$. Let

$$\epsilon := \left(\frac{tn - T(x)}{T(l) - T(x)} \right)^{1/m}.$$

For each $1 \leq i < j \leq N$, let

$$y_{ij} := x_{ij} + \epsilon(1 - x_{ij}).$$

Let $y_{ji} = y_{ij}$ and $y_{ii} = 0$. Then $y \in [0, 1]^n$, and by the inequality

$$\prod_{i=1}^r (a_i + b(1 - a_i)) \geq (1 - b^r) \prod_{i=1}^r a_i + b^r \quad (5.7)$$

that holds for any r and any $a_1, \dots, a_r, b \in [0, 1]$ (easy to prove by induction), we get

$$T(y) \geq (1 - \epsilon^m)T(x) + \epsilon^m T(l) = tn.$$

Thus, by the convexity of I_p ,

$$\begin{aligned} \phi_p(t) &\leq I_p(y) = I_p((1 - \epsilon)x + \epsilon l) \\ &\leq (1 - \epsilon)I_p(x) + \epsilon I_p(l) \\ &\leq I_p(x) + \epsilon n \log(1/p) = \phi_p(t - \delta) + \epsilon n \log(1/p). \end{aligned}$$

Since $T(x) \geq (t - \delta)n$,

$$\epsilon^m \leq \frac{tn - (t - \delta)n}{T(l) - (t - \delta)n} \leq \frac{\delta}{t_0 - t}.$$

This completes the proof of the lemma. \square

Lemma 5.8. *For any p and t ,*

$$\phi_p(t) \leq \frac{1}{2}(\lceil t^{1/k} N \rceil + k)^2 \log(1/p).$$

Proof. Let $r := \lceil t^{1/k} N \rceil + k$. Define $x \in [0, 1]^n$ as

$$x_{ij} := \begin{cases} 1 & \text{if } 1 \leq i < j \leq r, \\ p & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} T(x) &\geq \frac{1}{N^{k-2}} \sum_{q \in [r]^k} \prod_{\{l, l'\} \in E} x_{q_l q_{l'}} \\ &\geq \frac{r(r-1) \cdots (r-k+1)}{N^{k-2}} \geq tN^2 \geq tn, \end{aligned}$$

and since $I_p(p) = 0$,

$$I_p(x) = \sum_{i < j} I_p(x_{ij}) \leq \frac{1}{2} r^2 \log(1/p).$$

This proves the claim. \square

Proof of the upper bound in Theorem 1.2. The task now is to pull together all the information obtained above, for use in Theorem 1.1. As intended, we work with $f = T$. Take $t = \kappa p^m$ for some fixed $\kappa > 0$. Let δ and ϵ be two positive real numbers, both less than t , to be chosen later. Note that $\delta < t < \kappa p^{2m/k}$ since $t = \kappa p^m$ and $k > 2$. Assume that δ and ϵ are bigger than $N^{-1/2}$. Note that p is already assumed to be bigger than $N^{-1/2}$ in the statement of the theorem.

Recall that the indexing set for quantities like b_i and c_{ij} , instead of being $\{1, \dots, n\}$, is now $\{(i, j) : 1 \leq i < j \leq N\}$. For simplicity, we will write (ij) instead of (i, j) . Throughout, C will denote any constant that depends only on the graph H , the constant κ , and nothing else. From Lemma 5.1, we have the estimates

$$a \leq N^2, \quad b_{(ij)} \leq C,$$

and

$$c_{(ij)(i'j')} \leq \begin{cases} CN^{-1} & \text{if } |\{i, j, i', j'\}| = 2 \text{ or } 3, \\ CN^{-2} & \text{if } |\{i, j, i', j'\}| = 4. \end{cases}$$

Let $\theta := \delta^{-1}p^{2m/k}$. By Lemma 5.8,

$$K \leq Cp^{2m/k} \log N.$$

Using the above bounds, we get

$$\alpha \leq CN^2 \log N, \quad \beta_{(ij)} \leq C\theta \log N,$$

and

$$\gamma_{(ij)(i'j')} \leq \begin{cases} CN^{-1}\theta \log N & \text{if } |\{i, j, i', j'\}| = 2 \text{ or } 3, \\ CN^{-2}\delta^{-1}\theta \log N & \text{if } |\{i, j, i', j'\}| = 4. \end{cases}$$

Therefore, we have the estimates

$$\sum_{(ij)} \beta_{(ij)}^2 \leq CN^2 \theta^2 (\log N)^2, \quad \sum_{(ij)} b_{(ij)}^2 \leq CN^2,$$

and by Lemma 5.2,

$$\begin{aligned} \log |\mathcal{D}((\delta\epsilon)/(4K))| &\leq \frac{CN\theta^4}{\epsilon^4} \log \frac{CK}{\delta\epsilon} \\ &\leq \frac{CN\theta^4(\log N)^5}{\epsilon^4}. \end{aligned}$$

Combining the last three estimates, we see that the complexity term in Theorem 1.1 is bounded above by

$$CN^2 \epsilon \theta \log N + \frac{CN\theta^4(\log N)^5}{\epsilon^4}.$$

Taking $\epsilon = N^{-1/5}\theta^{3/5}(\log N)^{4/5}$, the above bound simplifies to

$$CN^{9/5}\theta^{8/5}(\log N)^{9/5}.$$

Next, note that by the bounds obtained above and the inequality $\delta > N^{-1/2}$,

$$\begin{aligned} \sum_{(ij)} \alpha \gamma_{(ij)(ij)} &\leq CN^3 \theta (\log N)^2, \\ \sum_{(ij), (i'j')} \alpha \gamma_{(ij)(i'j')}^2 &\leq CN^3 \theta^2 (\log N)^3, \\ \sum_{(ij), (i'j')} \beta_{(ij)} (\beta_{(i'j')} + 4) \gamma_{(ij)(i'j')} &\leq CN^2 \delta^{-1} \theta^3 (\log N)^3, \\ \left(\sum_{(ij)} \beta_{(ij)}^2 \right)^{1/2} \left(\sum_{(ij)} \gamma_{(ij)(ij)}^2 \right)^{1/2} &\leq CN \theta^2 (\log N)^2, \\ \sum_{(ij)} \gamma_{(ij)(ij)} &\leq CN \theta \log N. \end{aligned}$$

The above estimates show that the smoothness term in Theorem 1.1 is bounded above by a constant times

$$N^{3/2}\theta(\log N)^{3/2} + N\delta^{-1/2}\theta^{3/2}(\log N)^{3/2} + N\theta^2(\log N)^2.$$

Putting $\eta := p^{2m/k}$, we see that this is bounded by a constant times

$$N^{3/2}\delta^{-1}\eta(\log N)^{3/2} + N\delta^{-2}\eta^{3/2}(\log N)^2.$$

Since $\delta > N^{-1/2}$, we can further simplify this upper bound to

$$N^{3/2}\delta^{-1}\eta(\log N)^2.$$

Combining the bounds on the complexity term and the smoothness term, we get that

$$\begin{aligned} \log \mathbb{P}(T(Y) \geq tn) &\leq -\phi_p(t - \delta) + CN^{9/5}\delta^{-8/5}\eta^{8/5}(\log N)^{9/5} \\ &\quad + CN^{3/2}\delta^{-1}\eta(\log N)^2. \end{aligned}$$

By Lemma 5.7,

$$-\phi_p(t - \delta) \leq -\phi_p(t) + C\delta^{1/m}N^2 \log N.$$

Taking

$$\delta = N^{-m/(5+8m)}\eta^{8m/(5+8m)}(\log N)^{4m/(5+8m)}$$

gives

$$\begin{aligned} \log \mathbb{P}(T(Y) \geq tn) & \tag{5.8} \\ &\leq -\phi_p(t) + CN^{(9+16m)/(5+8m)}\eta^{8/(5+8m)}(\log N)^{(9+8m)/(5+8m)} \\ &\quad + CN^{(15+26m)/(10+16m)}\eta^{5/(5+8m)}(\log N)^{(10+12m)/(5+8m)}. \end{aligned}$$

Now note that since $p > N^{-1/2}$, therefore

$$\begin{aligned} \frac{N^{(9+16m)/(5+8m)}\eta^{8/(5+8m)}}{N^{(15+26m)/(10+16m)}\eta^{5/(5+8m)}} &= N^{(3+6m)/(10+16m)}p^{6m/k(5+8m)} \\ &\geq N^{(3+6m)/(10+16m)}N^{-3m/(5+8m)} \\ &= N^{3/(10+16m)}. \end{aligned}$$

This shows that the first term on the right-hand side in (5.8) dominates the second when N is sufficiently large. Therefore, when N is large enough,

$$\begin{aligned} \log \mathbb{P}(T(Y) \geq tn) & \\ &\leq -\phi_p(t) + CN^{(9+16m)/(5+8m)}p^{16m/k(5+8m)}(\log N)^{(9+8m)/(5+8m)}. \end{aligned}$$

Written differently, this is

$$\begin{aligned} &\frac{\phi_p(t)}{-\log \mathbb{P}(T(Y) \geq tn)} \\ &\leq 1 + \frac{CN^{(9+16m)/(5+8m)}p^{16m/k(5+8m)}(\log N)^{(9+8m)/(5+8m)}}{-\log \mathbb{P}(T(Y) \geq tn)}. \end{aligned}$$

By [19, Theorem 1.2 and Theorem 1.5],

$$-\log \mathbb{P}(T(Y) \geq tn) \geq CN^2p^\Delta, \tag{5.9}$$

where Δ is the maximum degree of H , provided that $p \geq N^{-1/\Delta}$ and N is sufficiently large. The lower bound on p is already assumed in the statement of the theorem. Therefore,

$$\begin{aligned} & \frac{\phi_p(t)}{-\log \mathbb{P}(T(Y) \geq tn)} \\ & \leq 1 + CN^{-1/(5+8m)} p^{-\Delta+16m/k(5+8m)} (\log N)^{(9+8m)/(5+8m)}. \end{aligned}$$

A minor verification using the assumption $p \geq N^{-1/4}$ shows that the ϵ and δ chosen above are both bigger than $N^{-1/2}$, as required. To complete the proof of the upper bound, notice that $\mathbb{E}(X)$ is asymptotic to p^m since $p \geq N^{-1/(m+3)}$. \square

Proof of the lower bound in Theorem 1.2. By Lemma 5.7, Lemma 5.1, and the lower bound in Theorem 1.1, ,

$$\log \mathbb{P}(T(Y) \geq tn) \geq -\phi_p(t) - CN^{-1/2m} N^2 \log N.$$

Therefore, again applying (5.9), we get

$$\frac{\phi_p(t)}{-\log \mathbb{P}(T(Y) \geq tn)} \geq 1 - CN^{-1/2m} p^{-\Delta} \log N.$$

This completes the proof of the lower bound. \square

6. PROOF OF THEOREM 1.4

In this section, all indices range over $\mathbb{Z}/n\mathbb{Z}$, and all additions and subtractions of indices are modulo n . As usual, C will denote any universal constant.

Let $Y = (Y_0, \dots, Y_{n-1})$ be a vector of i.i.d. *Bernoulli*(p) random variables. Define $f : [0, 1]^{\mathbb{Z}/n\mathbb{Z}} \rightarrow \mathbb{R}$ as

$$f(x) := \frac{1}{n} \sum_{i,j} x_i x_{i+j} x_{i+2j}.$$

Then

$$a := \|f\| \leq n. \tag{6.1}$$

Let $f_i := \partial f / \partial x_i$ and $f_{ij} := \partial^2 f / \partial x_i \partial x_j$. Then

$$f_i(x) = \frac{1}{n} \sum_j (x_{i+j} x_{i+2j} + x_{i-j} x_{i+j} + x_{i-2j} x_{i-j}).$$

From this expression, it is clear that

$$b_i := \|f_i\| \leq C, \quad c_{ij} := \|f_{ij}\| \leq \frac{C}{n}. \tag{6.2}$$

For each j , define the function $e_j : \mathbb{Z}/n\mathbb{Z} \rightarrow \mathbb{C}$ as

$$e_j(k) := \frac{1}{\sqrt{n}} e^{2\pi i j k / n},$$

where $i = \sqrt{-1}$. These functions form an orthonormal system, in the sense that

$$\sum_k e_j(k) \overline{e_{j'}(k)} = \delta_{j-j'},$$

where δ is the Kronecker delta function, that is,

$$\delta_j := \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{otherwise.} \end{cases}$$

For any $x \in \mathbb{R}^{\mathbb{Z}/n\mathbb{Z}}$, define its discrete Fourier transform $\hat{x} \in \mathbb{C}^n$ as

$$\hat{x}_j := \sum_k x_k e_j(k).$$

The orthonormality of the e_j 's implies the inversion formula

$$\begin{aligned} \sum_j \hat{x}_j \overline{e_k(j)} &= \sum_{j,l} x_l e_j(l) \overline{e_k(j)} \\ &= \sum_{j,l} x_l e_l(j) \overline{e_k(j)} = x_k. \end{aligned}$$

Moreover, it also implies the Plancherel identity

$$\begin{aligned} \sum_j |\hat{x}_j|^2 &= \sum_{j,k,l} x_k x_l e_j(k) \overline{e_j(l)} \\ &= \sum_{j,k,l} x_k x_l e_k(j) \overline{e_l(j)} = \sum_k x_k^2. \end{aligned}$$

Lemma 6.1. *For any $x, y \in [0, 1]^{\mathbb{Z}/n\mathbb{Z}}$,*

$$\sum_i (f_i(x) - f_i(y))^2 \leq C n^{1/2} \max_i |\hat{x}_i - \hat{y}_i|.$$

Proof. Note that for any x and y ,

$$\begin{aligned} &\sum_i (f_i(x) - f_i(y))^2 \tag{6.3} \\ &= \frac{1}{n^2} \sum_i \left(\sum_j (x_{i+j} x_{i+2j} + x_{i-j} x_{i+j} + x_{i-2j} x_{i-j} \right. \\ &\quad \left. - y_{i+j} y_{i+2j} - y_{i-j} y_{i+j} - y_{i-2j} y_{i-j}) \right)^2 \\ &= \frac{1}{n^2} \sum_{i,j,k} (x_{i+j} x_{i+2j} + x_{i-j} x_{i+j} + x_{i-2j} x_{i-j} \\ &\quad - y_{i+j} y_{i+2j} - y_{i-j} y_{i+j} - y_{i-2j} y_{i-j}) \\ &\quad \times (x_{i+k} x_{i+2k} + x_{i-k} x_{i+k} + x_{i-2k} x_{i-k} \\ &\quad - y_{i+k} y_{i+2k} - y_{i-k} y_{i+k} - y_{i-2k} y_{i-k}). \end{aligned}$$

Let us now expand out the product in the above expression. There will be 36 terms, 18 of which are positive and 18 are negative. The positive terms will be products of four x 's or four y 's, and the negative terms will be products of two x 's and two y 's. Match each positive term with a matching negative term. For example, match $x_{i+j} x_{i+2j} x_{i-k} x_{i+k}$ with $-x_{i+j} x_{i+2j} y_{i-k} y_{i+k}$. Summing over i ,

j and k for this particular pair, we get the expression

$$\begin{aligned} & \frac{1}{n^2} \sum_{i,j,k} (x_{i+j} x_{i+2j} x_{i-k} x_{i+k} - x_{i+j} x_{i+2j} y_{i-k} y_{i+k}) \\ &= \frac{1}{n^2} \sum_{i,j,k} (x_{i+j} x_{i+2j} (x_{i-k} - y_{i-k}) x_{i+k} - x_{i+j} x_{i+2j} y_{i-k} (x_{i+k} - y_{i+k})). \end{aligned} \quad (6.4)$$

Now consider the first term in the above expression. Let $z = x - y$. Then by the inversion formula,

$$\begin{aligned} & \frac{1}{n^2} \sum_{i,j,k} x_{i+j} x_{i+2j} x_{i+k} z_{i-k} \\ &= \frac{1}{n^2} \sum_{i,j,k} \sum_{a,b,c,d} \hat{x}_a \hat{x}_b \hat{x}_c \hat{z}_d \overline{e_{i+j}(a) e_{i+2j}(b) e_{i+k}(c) e_{i-k}(d)} \\ &= \frac{1}{n^{5/2}} \sum_{a,b,c,d} \sum_{i,j,k} \hat{x}_a \hat{x}_b \hat{x}_c \hat{z}_d \overline{e_{a+b+c+d}(i) e_{a+2b}(j) e_{c-d}(k)} \\ &= \frac{1}{n} \sum_{a,b,c,d} \hat{x}_a \hat{x}_b \hat{x}_c \hat{z}_d \delta_{a+b+c+d} \delta_{a+2b} \delta_{c-d} = \frac{1}{n} \sum_d \hat{x}_{-4d} \hat{x}_{2d} \hat{x}_d \hat{z}_d. \end{aligned}$$

By the Plancherel identity and the fact that $x \in [0, 1]^{\mathbb{Z}/n\mathbb{Z}}$, $\sum_j |\hat{x}_j|^2 \leq n$. In particular, $|\hat{x}_j| \leq \sqrt{n}$ for all j . Let $M := \max_i |\hat{z}_i|$. Using these observations and Hölder's inequality, we see that the above sum is bounded above by

$$\begin{aligned} & \frac{M}{n} \left(\sum_d |\hat{x}_{-4d}|^3 \sum_d |\hat{x}_{2d}|^3 \sum_d |\hat{x}_d|^3 \right)^{1/3} \\ & \leq \frac{CM}{n} \sum_d |\hat{x}_d|^3 \leq CM n^{1/2}. \end{aligned}$$

This is a bound on the first term in the right-hand side of (6.4). Similarly, it may be verified that the same bound holds for the second term in the right-hand side of (6.4), and also for all terms in the expansion of (6.3). This completes the proof of the lemma. \square

Lemma 6.2. *For the function f considered in this section, one can find sets $\mathcal{D}(\epsilon)$ satisfying (1.7) such that $|\mathcal{D}(\epsilon)| \leq C_1(n/\epsilon^2)^{C_2/\epsilon^4}$ where C_1 and C_2 are universal constants.*

Proof. Take any $\epsilon > 0$. Let $\gamma := c\epsilon^2\sqrt{n}$, where c is a universal constant that will be chosen later. Define a map $R : \mathbb{C}^n \rightarrow \mathbb{C}^n$ as follows: For each i , let the i th coordinate of $y = R(x)$ be the complex number closest to x_i whose real and imaginary parts are both integer multiples of γ . Clearly, $|x_i - y_i| \leq \gamma$. Moreover, if $|x_i| < \gamma/2$ then $y_i = 0$.

Let \mathcal{M} be the set of all \hat{x} as x ranges over $[0, 1]^n$. Take any $x \in [0, 1]^n$ and let $y := R(\hat{x})$. Let A be the set of all i such that $|\hat{x}_i| \geq \gamma/2$. Then $y_i = 0$ for each $i \notin A$. Given A , there are at most Cn/γ^2 possible values of each y_i , $i \in A$ since $|\hat{x}_i| \leq \sqrt{n}$. On the other hand by the Plancherel identity,

$$|A| \leq \frac{4}{\gamma^2} \sum_{i=0}^{n-1} |\hat{x}_i|^2 \leq \frac{4n}{\gamma^2},$$

implying that there are at most n^{4n/γ^2} possible candidates for the set A . Combining these observations, we realize that the number of possible values of y is at most

$$n^{4n/\gamma^2} (Cn/\gamma^2)^{4n/\gamma^2}.$$

This, therefore, is a bound on the size of $R(\mathcal{M})$.

Say that two points x and y in $[0, 1]^n$ are equivalent if $R(\hat{x}) = R(\hat{y})$. Clearly, this is an equivalence relation. Suppose that x and y are equivalent. Let $z = R(\hat{x}) = R(\hat{y})$. Then for each i ,

$$|\hat{x}_i - \hat{y}_i| \leq |\hat{x}_i - z_i| + |z_i - \hat{y}_i| \leq 2\gamma.$$

Construct the set B by choosing one x from each equivalence class. Then clearly

$$|B| \leq |R(\mathcal{M})| \leq n^{4n/\gamma^2} (Cn/\gamma^2)^{4n/\gamma^2}.$$

By the bounds obtained above and Lemma 6.1, for any $x \in [0, 1]^n$, there exists $y \in B$ such that

$$\sum_i (f_i(x) - f_i(y))^2 \leq Cn^{1/2} \max_i |\hat{x}_i - \hat{y}_i| \leq Cn^{1/2}\gamma.$$

The right-hand side is less than $\epsilon^2 n$ if the constant c in the definition of γ is chosen sufficiently small. Defining $\mathcal{D}(\epsilon)$ to be the set $\nabla f(B)$ completes the proof. \square

Lemma 6.3. *Let $\phi_p(t)$ be defined as in (1.5). Then for any $0 < \delta < t < 1$,*

$$\phi_p(t - \delta) \geq \phi_p(t) - \left(\frac{\delta}{1 - t} \right)^{1/3} n \log(1/p).$$

Proof. Take any $x \in [0, 1]^n$ such that $f(x) \geq (t - \delta)n$ and x minimizes $I_p(x)$ among all x satisfying this inequality. If $f(x) \geq tn$, then we immediately have $\phi_p(t) \leq I_p(x) = \phi_p(t - \delta)$, and there is nothing more to prove. So let us assume that $f(x) < tn$. Let

$$\epsilon := \left(\frac{tn - f(x)}{n - f(x)} \right)^{1/3}.$$

For each i , let

$$y_i := x_i + \epsilon(1 - x_i).$$

Then $y \in [0, 1]^n$, and by the inequality (5.7), we get

$$f(y) \geq (1 - \epsilon^3)f(x) + \epsilon^3 = tn.$$

Thus, by the convexity of I_p ,

$$\begin{aligned} \phi_p(t) &\leq I_p(y) \leq (1 - \epsilon)I_p(x) + \epsilon n \log(1/p) \\ &\leq I_p(x) + \epsilon n \log(1/p) = \phi_p(t - \delta) + \epsilon n \log(1/p). \end{aligned}$$

Since $f(x) \geq (t - \delta)n$,

$$\epsilon^3 \leq \frac{tn - (t - \delta)n}{n - (t - \delta)n} \leq \frac{\delta}{1 - t}.$$

This completes the proof of the lemma. \square

Lemma 6.4. *For any $p \geq n^{-1}$ and $t > 0$,*

$$\phi_p(t) \leq Ct^{1/2} n \log n.$$

Proof. Define $x \in [0, 1]^n$ as the vector whose first $3t^{1/2}n$ coordinates are equal to 1 and the rest are equal to p . Then

$$f(x) = \frac{1}{n} \sum_{i,j} x_i x_{i+j} x_{i+2j} \geq tn,$$

and $I_p(x) \leq Ct^{1/2}n \log n$. This proves the claim. \square

Lemma 6.5. *Suppose that $p \geq n^{-1/6}$. Then for any $\kappa > 1$,*

$$\mathbb{P}(f(Y) \geq \kappa p^3 n) \leq Ce^{-cnp^6}$$

where C and c depend only on κ .

Proof. Let $\kappa' := (1 + \kappa)/2$, so that $1 < \kappa' < \kappa$. It is easy to see that if n is sufficiently large (depending on κ), then $\mathbb{E}(f(Y)) \leq \kappa' p^3 n$. Again, (6.2) shows that $f(Y)$ changes at most by a bounded amount if one Y_i changes value. Therefore a straightforward application of Hoeffding's inequality [18] completes the proof of the lemma. \square

Proof of Theorem 1.4. Let $0 < \delta < t < 1$, where $t = \kappa p^3$ for some $\kappa > 1$ and δ is to be chosen later. Fix another small quantity ϵ , also to be chosen later. Assume that ϵ and δ are both bigger than $n^{-1/3}$. Already from the statement of the theorem, recall that $p \geq n^{-1/162}$.

Throughout this proof C will denote any constant that may depend only on κ . Let K, α, β_i and γ_{ij} be defined as in Theorem 1.1. By (6.1), (6.2), and Lemma 6.4,

$$K \leq Ct^{1/2} \log n, \quad \alpha \leq Cn \log n, \quad \beta_i \leq \frac{Ct^{1/2} \log n}{\delta}, \quad \gamma_{ij} \leq \frac{Ct^{1/2} \log n}{\delta^2 n}.$$

These imply the bounds

$$\begin{aligned} \sum_i \alpha \gamma_{ii} &\leq \frac{Ct^{1/2} n (\log n)^2}{\delta^2} \leq \frac{Ctn (\log n)^2}{\delta^{5/2}}, \\ \sum_i \beta_i^2 &\leq \frac{Ctn (\log n)^2}{\delta^2}, \quad \sum_{i,j} \alpha \gamma_{ij}^2 \leq \frac{Ctn (\log n)^3}{\delta^4}, \\ \sum_{i,j} \beta_i (\beta_j + 4) \gamma_{ij} &\leq \frac{Ct^{3/2} n (\log n)^3}{\delta^4}, \\ \sum_i \gamma_{ii}^2 &\leq \frac{Ct (\log n)^2}{\delta^4 n}, \quad \sum_i \gamma_{ii} \leq \frac{Ct^{1/2} \log n}{\delta^2}. \end{aligned}$$

Combining these, we see that the smoothness term is bounded by

$$Ct^{1/2} \delta^{-2} n^{1/2} (\log n)^{3/2} + Ct \delta^{-3} (\log n)^2 + Ct^{1/2} \delta^{-2} \log n.$$

Since $\delta > n^{-1/3}$, the above expression is bounded by

$$Ct^{1/2} \delta^{-2} n^{1/2} (\log n)^{3/2}. \tag{6.5}$$

On the other hand by Lemma 6.2 and the assumption that $\epsilon > n^{-1/3}$,

$$\log |\mathcal{D}(\epsilon)| \leq \frac{C \log(n/\epsilon^2)}{\epsilon^4} \leq \frac{C \log n}{\epsilon^4}.$$

Therefore the complexity term is bounded above by

$$C\epsilon t^{1/2}\delta^{-1}n \log n + \log\left(\frac{Ct^{1/2}\log n}{\delta\epsilon}\right) + \frac{Ct^2(\log n)^5}{\delta^4\epsilon^4}.$$

Choosing

$$\epsilon = t^{3/10}\delta^{-3/5}n^{-1/5}(\log n)^{4/5}$$

and recalling the assumed lower bounds on ϵ , δ and p , we see that the complexity term is bounded by

$$Ct^{4/5}\delta^{-8/5}n^{4/5}(\log n)^{9/5}. \quad (6.6)$$

By Theorem 1.1, Lemma 6.3, the bound (6.5) for the smoothness term, and the bound (6.6) for the complexity term, we get that

$$\begin{aligned} \log \mathbb{P}(f(Y) \geq tn) &\leq -\phi_p(t) + C\delta^{1/3}n \log n + Ct^{4/5}\delta^{-8/5}n^{4/5}(\log n)^{9/5} \\ &\quad + Ct^{1/2}\delta^{-2}n^{1/2}(\log n)^{3/2}. \end{aligned}$$

Now choose

$$\delta = t^{12/29}n^{-3/29}(\log n)^{12/29}.$$

Recalling that $t = \kappa p^3$, this gives

$$\begin{aligned} \log \mathbb{P}(f(Y) \geq tn) &\leq -\phi_p(t) + Cp^{12/29}n^{28/29}(\log n)^{33/29} \\ &\quad + Cp^{-57/58}n^{41/58}(\log n)^{39/58}. \end{aligned}$$

By the assumed lower bound on p , it is easy to see that the second term on the right dominates the third if n is large enough. Together with an application of Lemma 6.5, this completes the proof of the upper bound in Theorem 1.4. For the lower bound, recall that

$$\begin{aligned} \log \mathbb{P}(f(Y) \geq tn) &\geq -\phi_p(t + \delta_0) - \epsilon_0 n - \log 2 \\ &\geq -\phi_p(t) - C\delta_0^{1/3}n \log n - \epsilon_0 n, \end{aligned}$$

where

$$\epsilon_0 = \frac{1}{\sqrt{n}} \left(4 + \left| \log \frac{p}{1-p} \right| \right) \leq Cn^{-1/2} \log n$$

and

$$\delta_0 = \frac{2}{n} \left(\sum_{i=1}^n (ac_{ii} + b_i^2) \right)^{1/2} \leq Cn^{-1/2}.$$

An application of Lemma 6.5 completes the proof of the lower bound. \square

7. PROOF OF THEOREM 1.6

Let all notational conventions be the same as in Section 5. However, instead of a single H , consider l graphs H_1, \dots, H_l , and define T_1, \dots, T_l accordingly.

Throughout this section, C will denote any constant that may depend only on the graphs H_1, \dots, H_l . Define

$$f(x) := \beta_1 T_1(x) + \dots + \beta_l T_l(x).$$

Let $B := 1 + |\beta_1| + \dots + |\beta_l|$, as in the statement of the theorem. Let a , $b_{(ij)}$ and $c_{(ij)(i'j')}$ be as in Theorem 1.5. Clearly,

$$a \leq N^2 \sum_{r=1}^l |\beta_r| \leq CBN^2.$$

By Lemma 5.1, we get the estimates

$$b_{(ij)} \leq CB$$

and

$$c_{(ij)(i'j')} \leq \begin{cases} CBN^{-1} & \text{if } |\{i, j, i', j'\}| = 2 \text{ or } 3, \\ CBN^{-2} & \text{if } |\{i, j, i', j'\}| = 4. \end{cases}$$

Let $\mathcal{D}_1(\epsilon), \dots, \mathcal{D}_l(\epsilon)$ be the $\mathcal{D}(\epsilon)$'s for T_1, \dots, T_l . Define

$$\mathcal{D}(\epsilon) := \{\beta_1 d_1 + \dots + \beta_l d_l : d_r \in \mathcal{D}_r(\epsilon/\beta_r l), r = 1, \dots, l\}.$$

Clearly, for any $x \in [0, 1]^n$, there exists $d_1 \in \mathcal{D}_1(\epsilon/\beta_1 l), \dots, d_l \in \mathcal{D}_l(\epsilon/\beta_l l)$ such that

$$\sum_{i=1}^n (f_i(x) - (\beta_1 d_{1i} + \dots + \beta_l d_{li}))^2 \leq l \sum_{r=1}^l \sum_{i=1}^n \beta_r^2 (T_{ri}(x) - d_{ri})^2 \leq n\epsilon^2.$$

Therefore, $\mathcal{D}(\epsilon)$ satisfies the requirement of Theorem 1.5. Also,

$$|\mathcal{D}(\epsilon)| \leq \prod_{r=1}^l |\mathcal{D}_r(\epsilon/\beta_r l)|. \quad (7.1)$$

By the bounds on a , $b_{(ij)}$ and $c_{(ij)(i'j')}$ obtained above, the following estimates are easy:

$$\begin{aligned} \sum_{(ij)} a c_{(ij)(ij)} &\leq CB^2 N^3, \quad \sum_{(ij)} b_{(ij)}^2 \leq CB^2 N^2, \\ \sum_{(ij), (i'j')} a c_{(ij)(i'j')}^2 &\leq CB^3 N^3, \\ \sum_{(ij), (i'j')} b_{(ij)} (b_{(i'j')} + 4) c_{(ij)(i'j')} &\leq CB^3 N^2, \\ \sum_{(ij)} c_{(ij)(ij)}^2 &\leq CB^2, \quad \sum_{(ij)} c_{(ij)(ij)} \leq CBN. \end{aligned}$$

Combining these estimates, we see that the smoothness term is bounded by $CB^2 N^{3/2}$. Next, by (7.1) and Lemma 5.2,

$$\begin{aligned} \log |\mathcal{D}(\epsilon)| &\leq \sum_{r=1}^l \log |\mathcal{D}_r(\epsilon/\beta_r l)| \\ &\leq \frac{CB^4 N}{\epsilon^4} \log \frac{CB^4}{\epsilon^4}. \end{aligned}$$

Therefore, the complexity term (of Theorem 1.5) is bounded by

$$CBN^2 \epsilon + \frac{CB^4 N}{\epsilon^4} \log \frac{CB^4}{\epsilon^4}.$$

Taking

$$\epsilon = \left(\frac{B^3 \log N}{N} \right)^{1/5},$$

this gives the bound

$$CB^{8/5} N^{9/5} (\log N)^{1/5} \left(1 + \frac{\log B}{\log N} \right).$$

By Theorem 1.5, this completes the proof of the upper bound. The lower bound follows easily from Theorem 1.5 and the bound on $\sum c_{(ij)(ij)}$ obtained above. This finishes the proof of Theorem 1.6.

REFERENCES

- [1] BHAMIDI, S., BRESLER, G. and SLY, A. (2008). Mixing time of exponential random graphs. In *2008 IEEE 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 803–12.
- [2] BOLLOBÁS B. and RIORDAN, O. (2009). Metrics for sparse graphs. In S. Huczynska, J. D. Mitchell, and C. M. Roney-Dougal, eds., *Surveys in combinatorics 2009*, pp. 211–287, London Math. Soc. Lecture Note Ser. 365, Cambridge University Press, Cambridge.
- [3] BOLTHAUSEN, E., COMETS, F. and DEMBO, A. (2003). Large deviations for random matrices and random graphs. *Unpublished manuscript*.
- [4] BORGS, C., CHAYES, J. T., COHN, H. and ZHAO, Y. (2014). An L^p theory of sparse graph convergence I: limits, sparse random graph models, and power law distributions. *arXiv preprint*.
- [5] BORGS, C., CHAYES, J., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2008). Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, **219** no. 6, 1801–1851.
- [6] BORGS, C., CHAYES, J., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2012). Convergent sequences of dense graphs. II. Multiway cuts and statistical physics. *Ann. Math. (2)*, **176** no. 1, 151–219.
- [7] CARINCI, G., CHAZOTTES, J.-R., GIARDINÀ, C. and REDIG, F. (2013). Nonconventional averages along arithmetic progressions and lattice spin systems. *arXiv preprint*.
- [8] CHATTERJEE, S. (2005). *Concentration inequalities with exchangeable pairs*. Ph.D. thesis, Stanford University.
- [9] CHATTERJEE, S. (2007). Estimation in spin glasses: a first step. *Ann. Statist.*, **35** no. 5, 1931–1946.
- [10] CHATTERJEE, S. (2012). The missing log in large deviations for triangle counts. *Random Structures Algorithms*, **40** no. 4, 437–451.
- [11] CHATTERJEE, S. and DEY, P. S. (2010). Applications of Stein’s method for concentration inequalities. *Ann. Probab.*, **38** no. 6, 2443–2485.
- [12] CHATTERJEE, S. and DIACONIS, P. (2013). Estimating and understanding exponential random graph models. *Ann. Statist.*, **41** no. 5, 2428–2461.
- [13] CHATTERJEE, S. and VARADHAN, S. R. S. (2011). The large deviation principle for the Erdős-Rényi random graph. *European J. Combinatorics*, **32** no. 7, 1000–1017.
- [14] CHATTERJEE, S. and VARADHAN, S. R. S. (2012). Large deviations for random matrices. *Commun. Stoch. Anal.*, **6** no. 1, 1–13.
- [15] DEMARCO, B. and KAHN, J. (2012). Upper tails for triangles. *Random Structures Algorithms*, **40** no. 4, 452–459.
- [16] DEMARCO, B. and KAHN, J. (2012). Tight upper tail bounds for cliques. *Random Structures Algorithms*, **41** no. 4, 469–487.
- [17] DEMBO, A. and ZEITOUNI, O. (2010). *Large deviations techniques and applications*. Corrected reprint of the second (1998) edition. Springer-Verlag, Berlin.
- [18] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Stat. Assoc.* **58** 13–30.
- [19] JANSON, S., OLESZKIEWICZ, K. and RUCIŃSKI, A. (2004). Upper tails for subgraph counts in random graphs. *Israel J. Math.*, **142**, 61–92.
- [20] JANSON, S. and RUCIŃSKI, A. (2002). The infamous upper tail. Probabilistic methods in combinatorial optimization. *Random Structures Algorithms*, **20** no. 3, 317–342.
- [21] JANSON, S. and RUCIŃSKI, A. (2004). The deletion method for upper tail estimates. *Combinatorica*, **24** no. 4, 615–640.
- [22] KIFER, Y. (2010). Nonconventional limit theorems. *Prob. Theory Related Fields*, **148**, 71–106.
- [23] KIFER, Y. and VARADHAN, S. R. S. (2010). Nonconventional limit theorems in discrete and continuous time via martingales. *arXiv preprint*.
- [24] KIM, J. H. and VU, V. H. (2000). Concentration of multivariate polynomials and its applications. *Combinatorica*, **20** no. 3, 417–434.
- [25] KIM, J. H. and VU, V. H. (2004). Divide and conquer martingales and the number of triangles in a random graph. *Random Structures Algorithms*, **24** no. 2, 166–174.
- [26] LATAŁA, R. (1997). Estimation of moments of sums of independent real random variables. *Ann. Probab.*, **25** no. 3, 1502–1513.
- [27] LOVÁSZ, L. (2012). *Large networks and graph limits*. American Mathematical Society, Providence, RI.
- [28] LUBETZKY, E. and ZHAO, Y. (2012). On replica symmetry of large deviations in random graphs. *arXiv preprint*.

- [29] LUBETZKY E. and ZHAO, Y. (2014). On the variational problem for upper tails of triangle counts in sparse random graphs. *arXiv preprint*.
- [30] MCDIARMID, C. (1989). On the method of bounded differences. In *Surveys in combinatorics*, 148–188, London Math. Soc. Lecture Note Ser., 141, Cambridge Univ. Press, Cambridge.
- [31] SZEMERÉDI, E. (1978). Regular partitions of graphs. *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, pp. 399–401, Colloq. Internat. CNRS, 260, CNRS, Paris.
- [32] TALAGRAND, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.* **81** 73–205.
- [33] TAO, T. and VU, V. H. (2006). *Additive combinatorics*. Cambridge University Press.
- [34] VU, V. H. (2001). A large deviation result on the number of small subgraphs of a random graph. *Combin. Probab. Comput.*, **10** no. 1, 79–94.
- [35] VU, V. H. (2002). Concentration of non-Lipschitz functions and applications. Probabilistic methods in combinatorial optimization. *Random Structures Algorithms*, **20** no. 3, 262–316.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
SEQUOIA HALL, 390 SERRA MALL
STANFORD, CA 94305

souravc@stanford.edu
adembo@stanford.edu